3-7-2021

Gas Stations and the Wealth Divide: Analyzing Spatial Correlations Between Wealth and Fuel Branding

Jean-Carl Ende
Portland State University

Gas Stations and the Wealth Divide

Analyzing Spatial Correlations Between Wealth and Fuel Branding.

by

Jean-Carl Ende

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Urban Studies
in
Urban Studies

Thesis Committee:
Yu Xiao, Chair
Greg Schrock
Jiunn-Der Duh

Portland State University
2021

Abstract

The gasoline refining and sales industry has many peculiarities. One such oddity is a difference in sales, distribution and pricing between branded and unbranded gasolines. Although fuels leave the refinery a uniform commodity, branding determines entirely different marketing and pricing schemes, with entirely different volatility and risk premiums. In order to determine if this volatility is felt evenly across all wealth demographics, this study uses *t*-tests and CART models to analyze income, home value and other wealth-based indicators in the areas surrounding gas stations, to determine if there is a correlation between branding and wealth. The results show the wealth demographics surrounding branded stations are higher than around unbranded stations. I conclude that areas of higher wealth are more likely to have the presence of branded stations than unbranded, while areas of lower wealth have reasonable coverage by both branded and unbranded.

Table of Contents

List of Tables

List of Figures

Introduction

The retail gasoline market has many facets and hidden intricacies, two of which are branding and pricing. Branding dictates which marketing scheme is used, which in turn dictates the price setting mechanism. Branded fuels are sold through long-term contracts, but unbranded fuels are sold through competitive bidding wars. This implies that prices at unbranded stations are less likely to be monopolistically set, but with that resulting competition comes price uncertainty and volatility. If a shortage is expected, prices may suddenly rise. Similarly, if a shortage is realized, unbranded retailers may be cut off so that refiners can fulfil long term branded retail contracts. Thus, volatility and uncertainty are passed on to those consumers who are more likely to frequent unbranded stations, rather than those who frequent branded stations. But how might those consumers best be identified?

In economics, general assumptions are made where prior theory exists and empirics are difficult to obtain or assess. In this case, an assumption might be made that, since unbranded fuel is generally a little cheaper than branded, unbranded stations are more likely to be located in areas with lower-income residents, while branded stations – a little more expensive – are more likely to be located in higher wealth areas. If this is verified to be true, it would imply that lower-income households are more likely to feel the pinch through fuel pricing volatility and occasional outages, while wealthy households would be subjected to a less competitive system of pricing. This study is intended to be a reality check on these assumptions by exploring the empirics of station locations and their surrounding demographics in the Portland Metropolitan Area.

Section 1: Background - How Gasoline Branding Works

While it may not be self-evident, all gasoline is in fact created equal. That is to say, gasoline is a commodity: given a host of specifications which depend on state or regional regulation, the “regular” and “premium” gasolines that leave one refinery are identical to the “regular” and “premium” fuels leaving competitors’ refineries. This allows companies to store and transport fuel in the same tanks and pipelines, as well as buy, sell and trade gasoline in real time, should one refiner find that they are suddenly in need of more supply to fulfill contracts, or are in a supply glut. This system keeps the entire network of stations supplied with fuel, and free from shortages. (AAA, 2016)

However, that does not mean that retail fuel stations are all the same. On the West Coast, the most prominent “branded” stations are Chevron, Shell, BP, Exxon and Phillips 66 (labeled as “76” stations, formerly Union 76). These are stations that are owned by the refinery and share their brand name, or contract with them so that they can be a part of their distribution and pricing network. Unbranded are typically stations like Arco, Costco, Space Age, Safeway and other small corner markets or off-brand stations that don't share a brand name with a major refinery, and thus aren't required to exclusively sell gasoline under a branded marketing contract. (Kendrick Oil, 2017) Below, Figure 1, produced by the Energy Information Administration (EIA), details how branded and unbranded fuels go from refinery to market:

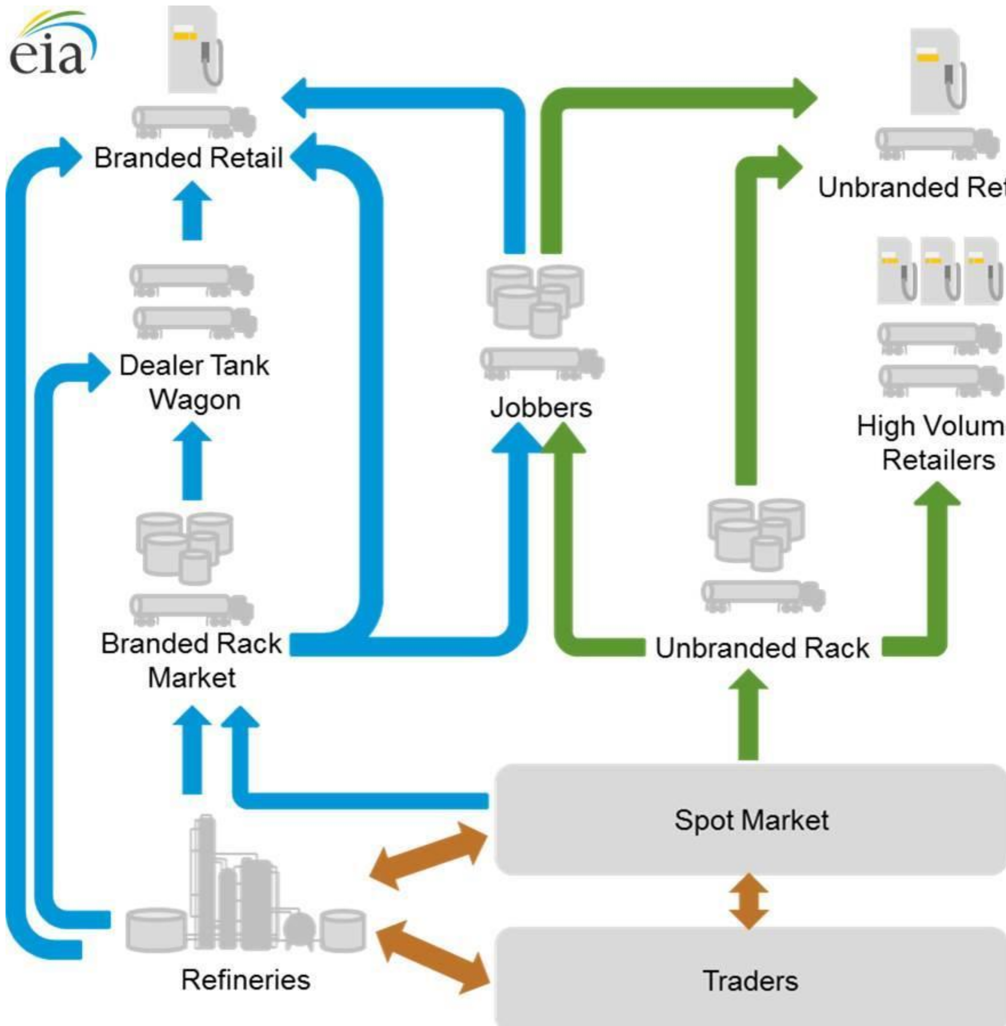

Figure 1 - Refinery to Retail: Diagram depicting how gasoline and gasoline pricing reaches the retail markets via separate branded and unbranded avenues. Source, Energy Information Administration. (EIA, 2015, p. 19)

As shown, there is a split between "branded" and "unbranded" marketing schemes that is inherent to the gasoline wholesale and retail industries. In general, branded stations are supplied directly from the refineries, through the branded rack, or through jobbers. Jobbers are speculators and middlemen who obtain medium sized lots of fuel from the wholesale rack and sell truckloads to smaller retailers or larger-scale consumers, like farms and fleets. (EIA, 2015)

Unbranded stations purchase fuel supplied through the unbranded rack, which is supplied by and priced based on the spot market. (OPIS, 2020) The spot replacement market is a contracts and futures market, much like Wall Street, but for refined petroleum products and components. Refiners, wholesalers, jobbers, and even speculative buyers with no intention of receiving the product, all interact to buy, sell and trade fuel contracts. The prices are recorded daily by price reporting agencies – such as the Oil Price Information Service (OPIS) – who receive trading reports detailing the transactions of the day, and their prices. The next day, trading resumes with this reporting as the reference price.

Because the spot market is a competitive bidding open-outcry marketplace, supposedly free from contract premiums and monopolistic market powers, unbranded fuel tends to be somewhat less expensive than branded. While the price of unbranded fuel reported by OPIS may be used in the pricing mechanism of the branded market, there can be a significant delay between spot market price setting and the prices paid by refinery-owned or franchised wholesalers, due to the nature of long-term supply contracts. Because of this bid/offer system, and the delay in branded price setting, the wholesale prices of unbranded fuels are often more volatile than branded. (NACS, 2020)

Ultimately, the fuel that comes from the refinery, and is received at both branded and unbranded racks, is so identical that it is all transported in the same pipes and tanks. It is at the rack that ethanol is added along with proprietary detergents and chemicals for branded stations, (for instance, Chevron's Techron). Unbranded fuels do not have proprietary detergents and chemicals added. (Kendrick Oil, 2017) There is much debate as to whether these detergents and chemicals change the performance of branded gasoline, calling into question whether the only real differences are based solely on consumer perception, (AAA, 2016) but that course of analysis is not within the scope of this study.

It is important to emphasize that each major brand of branded gas has its own proprietary pricing methodology that determines the wholesale and retail prices of its branded stations. While the overall price of gas that leaves the refinery is highly dependent on the price of crude oil, branded stations are largely insulated from short-run speculative activities due to the nature of long-term contracting. In a broader sense, this is concerning because the branded market exhibits a lack of competition and instead, a type of monopolistic "price-setting" market power. Meanwhile, unbranded gasoline is priced exclusively through the spot market, which is subject to the whims of speculators who place bids based on the economic fundamentals of supply, demand and profitability; a far more competitive system, but also more volatile. From the same EIA report mentioned above:

> *There are about 15 to 20 participants in the West Coast spot market, including refiners that buy and sell products to balance refinery production and sales commitments, trading companies that are in the business of buying and selling gasoline but that typically have no presence*

> *in wholesale or retail gasoline markets, brokers with market knowledge and understanding that identify buyers and sellers and arrange deals, and independent retail marketers that move large volumes of gasoline through their own retail outlets. Prices in the spot market move with perceived changes in refinery supply and demand.* (EIA, 2015, p. 18)

While unbranded stations have the freedom to purchase products from whomever they want, there is usually only one unbranded wholesale price, and it may change very rapidly due to free market competition. In addition to this price volatility, if refineries are facing a sudden shortage and need to carefully budget supplies in order to fulfill in-house contracts, major spot market sellers may be entirely cut off from supply for a short time. This may cause what is known as a "price inversion", where speculators in the spot market react to the supply shortage by fighting over the remaining gasoline and bidding up the price until it is higher than the wholesale price of branded gas. (Kendrick Oil, 2017) Thus, wholesale consumers of unbranded gasoline face what economists call "price uncertainty", while wholesale customers of branded gasoline face a monopolistic "price-setting" market.

Section 2: Research questions

This paper seeks to reveal connections between fuel branding and wealth demographics within the Portland, Oregon Metropolitan Urban Growth Boundaries (Portland Metro UGB). Since station branding can be understood as a proxy for price volatility and market competition, can wealth indicators surrounding a given station be used to predict whether that station is likely to be branded or unbranded? Answering this research question can help us understand questions such as, “Is wealth in the areas around branded stations higher than around unbranded?”, and “Do areas of differing wealth face different fuel price volatility and competition?”

Section 3: Literature review

Literature on the subject is largely absent, but a few authors have studied the differences in branded and unbranded fuel markets. As early as 1953, in a work titled *Price Influence of Unbranded Gasoline,* Vernon T. Clover studied the differences between what he labels “Standard” and “Independent” stations. Surveying the gas stations of four (4) separate, yet justifiably similar cities in Texas, he gathered data on branding, appearance, prices and services offered. His research focused on whether core economic principles were in fact true, concluding that in many ways they are not. His research uncovered the fact that, to some extent, the gasoline market seems to defy the assumptions of a competitive marketplace, as well as the price-changing influence of supply and demand. Clover found that prices in the gasoline market are, in his words, “flexible”, or rather, they lack uniformity. He suggested that while independent stations charge less than standard stations, it is only by about 2-4 cents (in 1950s currency valuation), and in a uniform way. He found there was not a statistically significant correlation between a greater number of independent stations and more competitive pricing. To him, this meant that independent stations priced their gas based on the price of the nearby standard stations, not based on competitive market fundamentals within the unbranded market. His findings suggest that even back in the 1950s there was a lack of competition in the market. However, his work was centered on the price influence of market fundamentals, not the location and branding of stations relative to urban wealth. (Clover, 1953)

In 2008, economists Doyle and Samphantharak used location data to analyze purchasing decisions at gas stations near state borders, studying an individual's willingness to drive in order to save money on gas. They found that in states that had recently imposed a gas tax, stations within five miles of the border saw sales fall, while nearby stations in a neighboring state that did not impose the same tax rose. This suggests that some people are willing to drive to avoid paying more, but that convenience and the amount of the price difference are big factors as well. In some more extreme cases, where metropolitan areas are immediately adjacent to state borders, station owners were forced to cut prices to more closely match the untaxed stations, in order to win back customers. Doyle and Samphantharak also found that in most cases, states that have significant border populations will increase taxes in tandem to avoid this sort of tax-shirking problem. This is, however, about the behavior altering effects of taxation and the decisions people make, given a new constraint. (Doyle & Samphantharak, 2008)

Recently publishing their work in 2020, French researchers Bergeaud and Raimbault also studied the spatial variability of fuel prices by generating a unique data set and analyses. They modeled gas prices over a two (2) month period across the entire United States, finding that the main drivers of fuel price were already well-known, such as crude oil prices, regional distribution challenges, and state and local taxes. But they also found many local drivers of price variance stemmed from socio-economic processes, such as wage, income, population density and cultural differences. Their study aided in refining this paper's modeling techniques, but did not seek the same information or conclusions. (Bergeaud & Raimbault, 2020)

Because the literature is relatively sparse on this exact topic, or is focused on analysis of fuel prices, this paper fills a gap regarding fuel branding, distribution, and the socio-economic variance of sudden shocks in price and supply. In the conclusions and discussions, I also linked the findings to other ways of thinking about spatial distribution in metropolitan areas such as central place theory, the study of gentrification and generally understanding how neighborhoods and cities change and evolve over time.

Section 4: Data & Analytical Methods

In this section, I discuss the data sources and analytical methods. The subsections are as follows; Section 4.1 explains how the initial dataset of gas station locations were obtained. Section 4.2 explains how location and branding data was cleaned and verified. Section 4.3 explains how service areas were generated, using station locations. Section 4.4 explains how wealth indicators were generated, using service areas.
Finally, in Section 4.5, analytical methods are discussed.

4.1 Gas Station Locations

Initial gasoline station data came from ReferenceUSA (very recently changed to Reference Solutions). This website contains comprehensive lists of public and semi-private businesses, along with certain information and other attributes, where possible. Found under “Major Industry Group”, “Retail Trade”, and #55, “Automotive Dealers and Service Stations”, data set #5541, “Gasoline Service Stations” forms the basis of this study’s location database. (ReferenceUSA, 2020)

Station location data consists of vital information, such as name, address, and GPS coordinates denoted in latitude and longitude. Attached to each location is attribute data such as owner's name, manager, contact information, slogan, online media links and a number of other details. There is also information about conjoined retail establishments, like convenience stores or restaurants. This could help identify characteristics that may be useful for future analyses, but the data is semi-inconsistent, with numerous gaps. All locations within the three counties that span the Portland Metro area (Multnomah, Washington, and Clackamas) were initially queried with the understanding that there would be significant trimming to restrict to the Portland Metro UGB.

4.2 Verification of Gas Station Data

For this study, the most important data was simply the list of location addresses and their latitude - longitude coordinates. To begin verification, location data was loaded into ArcGIS. This significantly aided the process by giving a visual representation of the data that could be cross-checked against other maps with locations. Locations were then trimmed to only those that fall within the UGB of the Portland metro area. GIS shapefiles of this boundary are available at the Oregon Spatial Data Library. (Spatial Data Library, 2014)

To verify the gas station locations and their brands, ArcGIS locations were intensively cross-checked against a Google Maps “street view” search for the term “gas station”. This can be considered “virtual ground-truthing”, as it uses imagery from firsthand observations of the physical location in question, as opposed to firsthand observations, themselves. In other words, I didn’t go to each location myself, but someone was physically present at the location to take the picture. Thus, I only “virtually” ground-truthed the location data. This is a reasonable method because Google’s 360-degree panoramic imaging feature allows an objective view of any physical location. Since the vast majority of Google Street View images were taken as recently as 2019 or later, a nearly current-day verification of every square meter of the Portland Metro area was possible. (Google, 2020)

For any Google results that were not clear, or locations that seemed to be in a state of flux, business phone listings were referenced and used to verify their current status. For those phone numbers that were inaccurate or outdated, nearby businesses were contacted which were happy to confirm the operation and brand of the gas station across the street. In one particular case this was critical: a station that might have been removed as non-operational was found to not only be functioning and operational, but it had switched from unbranded to branded within the year. Its listed phone number was no longer functioning, so speaking to the manager at the station across the street yielded valuable information about the station's history and current status.

Since the latitude and longitude coordinates were also available in Google Maps, each location's address could be cross-verified with its GPS coordinates. This is important because more than one observation had coordinates that didn't quite match its address, and needed to be corrected.

Road by road, neighborhood by neighborhood, inspections were conducted using this technique to not only verify each individual observation in the data set, but to meticulously inspect the entire Portland Metro area for stations that were skipped over and not recorded in the initial list. More than a few Shell stations, or AM/PM stations were simply not on the list, but were without question in operation and pumping gas, and had been for quite some time in the past. So, to consider the list comprehensive, they were included.

Likewise, a number of stations had been converted into a repair shop or a coffee shop, and no longer possessed pumps, even though they were still listed as fueling stations by Google. This meticulous row-by-row verification method also revealed a number of observations in the ReferenceUSA list that did not coincide with gas stations, i.e., Plaid Pantry, IKEA, Walgreens, McDonalds, Providence Medical Center, Midland library, City Hall, etc. These were removed. A number of “cardlock” stations were also removed. Cardlock stations are business account stations for fleets and service vehicles that operate on special credit cards. They are not used by the general public and are often located in somewhat more industrial areas. Any observations removed from the main data set were placed in a separate spreadsheet so as to not destroy data.

While many attributes were included in the data set, station branding was not indicated. However, during the cross-verification of location data against Google business listings, and signage in the Google street view image, branding was able to be determined and the information added to the database. It should be noted that there are only three (3) major branded brands in the Portland Metro area; Chevron, Shell, and Phillips 66, aka 76 (previously Union 76, or Unocal 76). Each of these brands own a refinery in Washington State, and a distribution and retail sales network throughout the northwest. They also sign contracts with local franchisees who want to ensure their supply and advertising network. All other stations are supplied as unbranded, meaning they can buy from any source, the branded wholesalers, or other refiners, such as Marathon that also has a refinery in Anacortes, Washington, and is known for selling unbranded fuels. (Tesoro, 2006, p. 10)

There are also six (6) Exxon stations which are yet to be determined if they are branded or not. Exxon is usually considered branded, but without an established system of refining and wholesale supply, and with such meager station numbers, it is hard to be sure if they are considered branded. In this case, it may be that they are supplied through a contract with one of the branded suppliers, but are able to retain their own brand signage. Alternatively, it may be that they want to begin establishing a presence in the area and starting out unbranded gets their name out there. Whatever the reason, their locations are relatively inconsequential, so they were labeled as branded, since that is how Exxon is known nationwide. (Exxon, 2020)

The original data set was very rough. It contained just over eight hundred (800) observations, but many were invalid. After the first steps of removing duplicates, non-stations, and those located far from Portland, the number of observations shrank significantly to just over four hundred (400). In the end, after meticulous inspection of the data, a total of just under three hundred (300) locations were verified. Figure 2 below is a map showing their locations and the Portland Metro UGB:

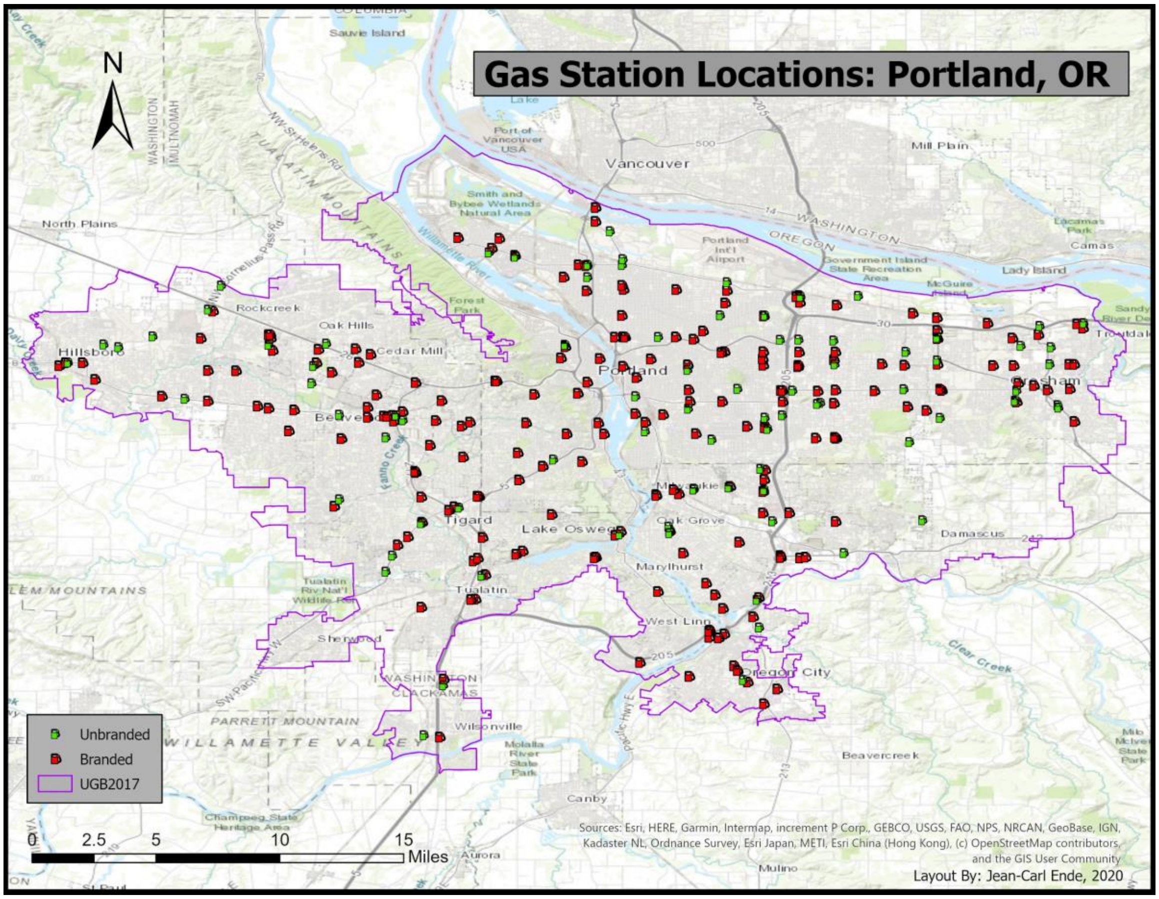

Figure 2 - Gas Stations Locations: Service Stations within the Portland Metro UGB. Branded stations are in red, unbranded in green.

In Figure 2, branded (red) and unbranded (green) stations can be seen scattered across the Portland metro area. In some areas the locations appear random, while in others they seem to follow major arterial transportation thoroughfares. Some large gaps also exist where there are protected natural areas, large agricultural plots of land or terrain unsuitable for habitation. Because of the arrangement of stations along major roads, and not necessarily in all locations, it was important to establish local service areas from which to gauge the demographics of each station.

4.3 Service Areas

In order to understand the demographics of those most likely to visit a given station, service areas were generated. In other industries, service areas are generally used for a number of purposes, like businesses trying to understand the locations of the closest competition or their nearest suppliers. But, unlike determining a simple one-mile radius from a given point, service areas are generated by following the roadways and traffic patterns, giving a more accurate driving distance, particularly where there are many waterways or other natural barriers that make it impossible to drive past. In this way they can be used for delivery route planning, or service scheduling. In most of these cases, a "transit network" data set is necessary for generating these service areas. Unfortunately, the transit network data available from the State of Oregon is limited. (Spatial Data Library, 2019) However, the Environmental Systems Research Institute (ESRI) maintains a database that can be queried through their GIS software program, ArcGIS.

Using ArcGIS's "Generate Service Areas" tool, service areas with a number of specific attributes were generated. (ESRI, 2020) For this particular study, three separate queries were run with "Break Values" set at 1, 2 and 3. "Break Units" were set to "Miles" instead of "Minutes", the "Travel Direction" was set to "Towards Facility" instead of "Away from Facility", and all road and surface type restrictions were lifted. Restrictions were lifted because roads under construction, gated, private, unpaved and dead-end roads, all may have people living on them with recorded household wealth. The intention was to not exclude those households simply because they are not accessible by the public. One restriction was kept: "Driving an Automobile". The "Impedance" and "Distance

Impedance" were changed to "Miles". All other values were left in their default position. Figure 3 below is a closeup example of two (2) stations with their service areas, a branded and an unbranded:

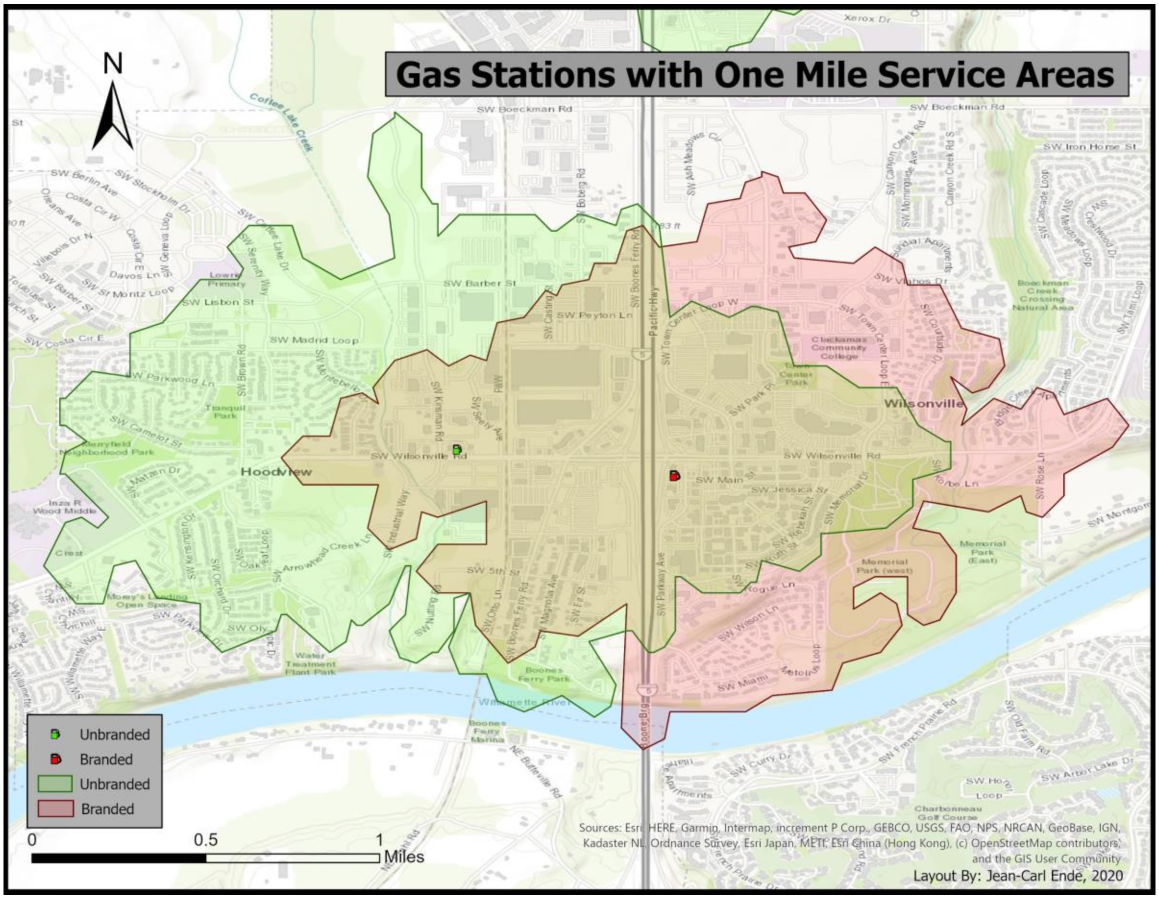

Figure 3 - Individual Service Areas: Unbranded (green) and branded (red) service areas follow the transit network, creating a driving distance that is at maximum one (1) mile.

You can see from Figure 3 that by giving a buffer zone around the traffic network, the homes and demographics can be captured in a greater area than the roadway itself. In addition, the river acts as a natural barrier so that houses in another region were not included in the analysis

4.4 Wealth Indicators

ESRI's Community Analyst provides census data which comes from official sources, including the US Census and ACS surveys. Data can include population, income, employment, health, density, race and many other demographic indicators. Service areas were loaded into Community Analyst and a "Comparison Report" was run. (ESRI, 2020) A comparison report queries ESRI's database of location-based information and finds an average value for a given shapefile. It is this list of demographic information that forms the backbone of the analysis. Table 1 below includes the main list of demographic and wealth indicators sourced for this study, along with their year and the variable name used in various outputs:

Table 1 - Wealth Indicators

ESRI Community Analyst - Wealth Indicators		
Year	**Wealth Indicator**	**Variable**
2020	Median Household Income	medincome
2020	Median Disposable Income	dispincome
2018	Households Receiving Food Stamps/SNAP	SNAPperc
2020	Median Net Worth	mednetworth
2020	Median Home Value	medhomevalue
2020	Wealth Index	wealthindex
2020	Per Capita Income	percapinc
2020	Average Household Income	avghouseinc

The indicator "Wealth Index" is specialty data generated by ESRI using census data. Below is an excerpt from ESRI's website describing their approach to the Wealth Index indicator:

> *The wealth index is designed not to evaluate worth, but rather to capture the standard of living and financial stability of area households. Esri's wealth index represents a scale of an area's wealth relative to the national level. An index of 100 represents wealth on par with the national average. An area with a wealth index below 100 has lower than average wealth, while an index above 100 identifies areas with above average wealth.* (ESRI, 2020)

Rather than simply reporting a static numerical representation of wealth, which may not make any sense in different locations, the Wealth Index gives a figure that is comparable in different locations and economic situations. (Esri, 2020)

Upon inspecting the census data for problematic observations, it became apparent that there were nine (9) outlier stations that did not fit with the rest of the data. Stations with a total population below three hundred fifty (350) were identified as problematic and removed. While it may seem like a fair number of residents, this is an inordinately low number, considering most service areas represent well over one thousand (1000), and many are over ten thousand (10,000).

The reasoning behind this removal was that the practice of generating service areas and determining a demographic value for a given station assumes that the people within that service area somehow represent the people that might frequent the station. But those stations that have such a low population nearby are not a typical neighborhood gas station. In this case, after individual verification, it turns out each fall into one of two types of locations; industrial/commercial areas with large stores, warehouses and industrial parks, or rural highway intersections in between one township and another. In both cases, the majority of the customers of the station are not choosing that station because they live nearby, (thus, their home and income demographics are not represented in the service area); they are choosing it because it is on their way between home and another place. It is convenient. Similarly, those station owners are not basing their branding decisions on the residents that live nearby, but instead, on the traffic that flows past. This makes these stations too different to be suitable for inclusion in the analysis. After this final removal of stations, there were two hundred eighty-eight (288) remaining. Table 2 contains a frequency chart of stations by brand and branding.

Table 2 – Brand Frequency Table

Frequency Table		
Branding	**Brand**	**Freq.**
Branded	76	58
Branded	Chevron	81
Branded	Exxon	6
Branded	Shell	61
Unbranded	Unbranded	82
Branded	Subtotal	206
Grand Total		**288**

Table 2 shows who the major branded companies are, and the number of stations that are contracted through each brand. It also shows the number of branded and unbranded stations, and the grand total.

4.5 Analytical Methods

To undertake this study, gas stations within the Portland Metro UGB were compared based on the wealth indicators of those consumers most likely to frequent the station. While it is highly presumptive to assume that a station's customer base is comprised solely of the residents within a certain proximity, it is reasonable to assume that a person sitting in their home, thinking about where to get gas the next time they leave, will at least consider those stations that are closest to their home. Likewise, any station owner trying to decide whether to maintain an unbranded station, or to remodel and seek a branding contract, is likely to look at the surrounding area and its wealth demographics, among other factors. To undertake this, each location is associated with a drive-distance service area of one (1) mile. Service areas at two (2) and three (3) mile drive distances were also constructed, but were determined to be too overlapping, and thus too collinear to be of any analytical value.

Census data for each of the service areas was gathered using ESRI's Community Analyst database queries. Two-sample *t*-tests (with presumed unequal variances) were conducted on each variable of the census data in order to analyze the fundamental differences between the service areas around branded and unbranded stations. These gave insight into how wealth indicators differ, on average.

Classification And Regression Tree (CART) models were then developed, using the census data. CART models construct a tree of regression "decisions" that split the data. As Diego Lopez Yse puts it, "CART algorithm uses a metric called Gini Impurity to create decision points for classification tasks. Gini Impurity gives an idea of how fine a split is (a measure of a node's "purity")...". (Yse, 2019) This technique allows highly influential observations to be separated out, successively, telling a story about the data. Each "branch" in the chart represents a different partition in the data where one variable, in a particular range of values, produces a very "pure" regression, or where the two resulting individual regressions have a better fit than the combined data. The output also suggests that the first branch has the highest impact on the dependent variable, and each successive branch represents the next most influential variable and break point.

The end results are a series of "leaf nodes". Each leaf node represents a simple analysis of the observations contained within that particular partition. For binary and categorical dependent variables (in this case, branded (1) and unbranded (0)), a "winner" and a propensity score are determined, which tells what the likelihood is that an observation will be among the "winner" group. A percentage of the total observations that are contained within that partition are also reported for each node. This results in a series of stories about each resulting cluster (or leaf node) within the data.

Section 5: Findings

To reiterate the purpose of the study, branded and unbranded stations sell nearly identical gasoline that is run through different marketing, pricing and distribution systems, with unbranded having a more volatile price structure and the potential to run out. An economic assumption might suggest that unbranded stations are predominantly in lower wealth areas. If, in general, stations are randomly distributed, but there are fewer unbranded stations located in areas of higher wealth, then it can be suggested that low- and middle-income households are more likely to face price volatility and, in extreme circumstances, gasoline shortages. These results use empirics to support the assumption that unbranded stations are predominantly excluded from areas of higher wealth, by comparing the demographics around branded gas stations to the demographics around unbranded stations.

The subsections are as follows: Section 5.1 explores the findings from a simple visual inspection of the data. Section 5.2 explores the results of a correlation matrix constructed with all variables. Section 5.3 explores the results of *t*-tests conducted on each variable. Section 5.4 explores the results of classification trees constructed from the data.

5.1 Descriptive Analysis

The first observations about the data can be made when visualizing the station service areas with mapping software. Figure 4 shows unbranded stations (green) layered on top of branded (red).

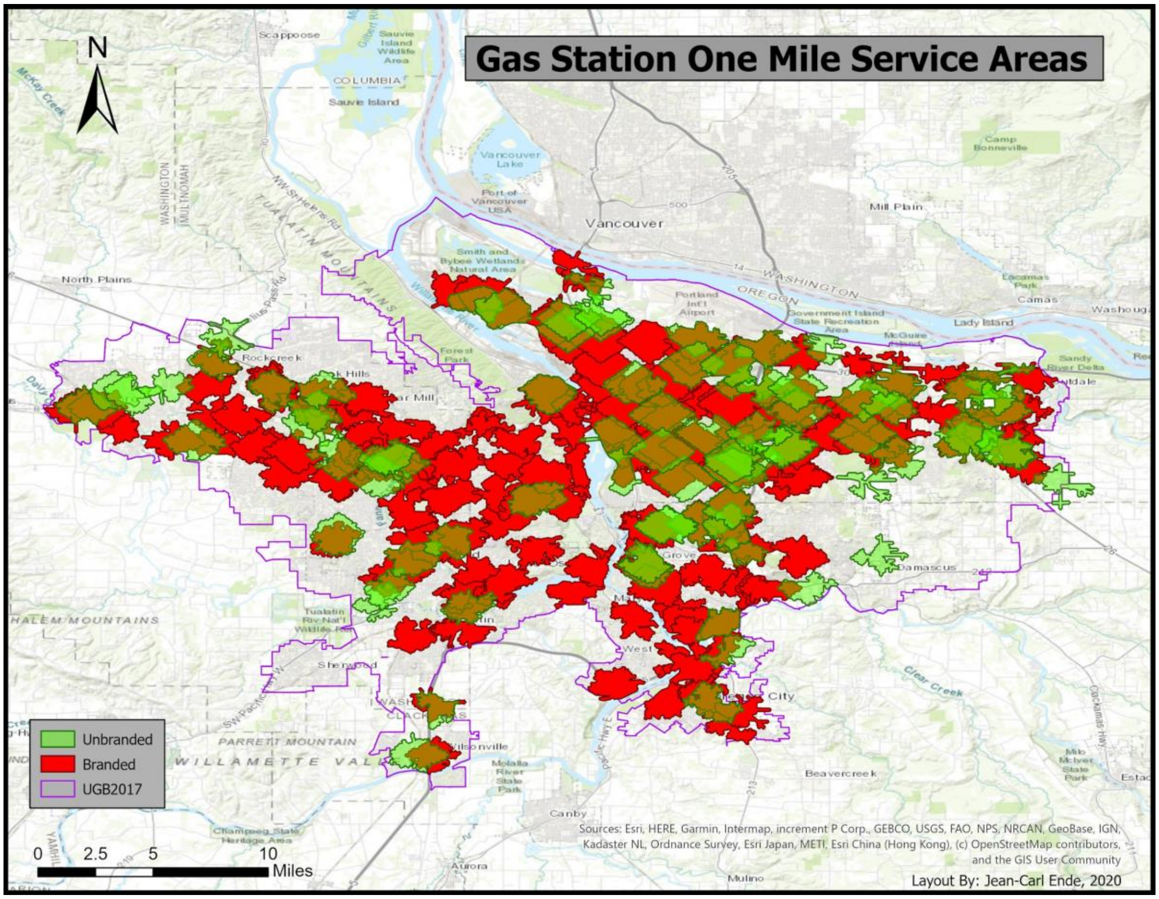

Figure 4 - All Service Areas: Unbranded gas station 1-mile service areas (green), and branded gas station 1-mile service areas (red), with the Portland Metro UGB (purple).

Immediately, it becomes apparent that unbranded stations do not have an even spread across the Portland Metro UGB in the same way as branded. On the right side of the map, east of the Willamette River, which roughly bisects the map in the middle, there is a fairly good coverage of green service areas. Red can still be seen through the gaps, but the coverage of green is fairly even and uniform. However, just to the left of center, and in the southern areas, unbranded become a bit sparse. Red can be seen through the green in a lot of places, and some areas seem completely devoid of unbranded stations. For those familiar with the Portland Area, those are townships named Lake Oswego, West Linn, Oregon City, Gladstone and generally the Downtown and West Hills/Hillsdale areas. These communities are where some of the most expensive homes and the highest concentration of personal wealth are located.

Figure 5 was created using ESRI's Community Analyst to illustrate the areas of higher and lower income around the Portland metro area. The areas in Figure 4, identified as those lacking unbranded stations, stand out as similar to those that have higher incomes in Figure 5.

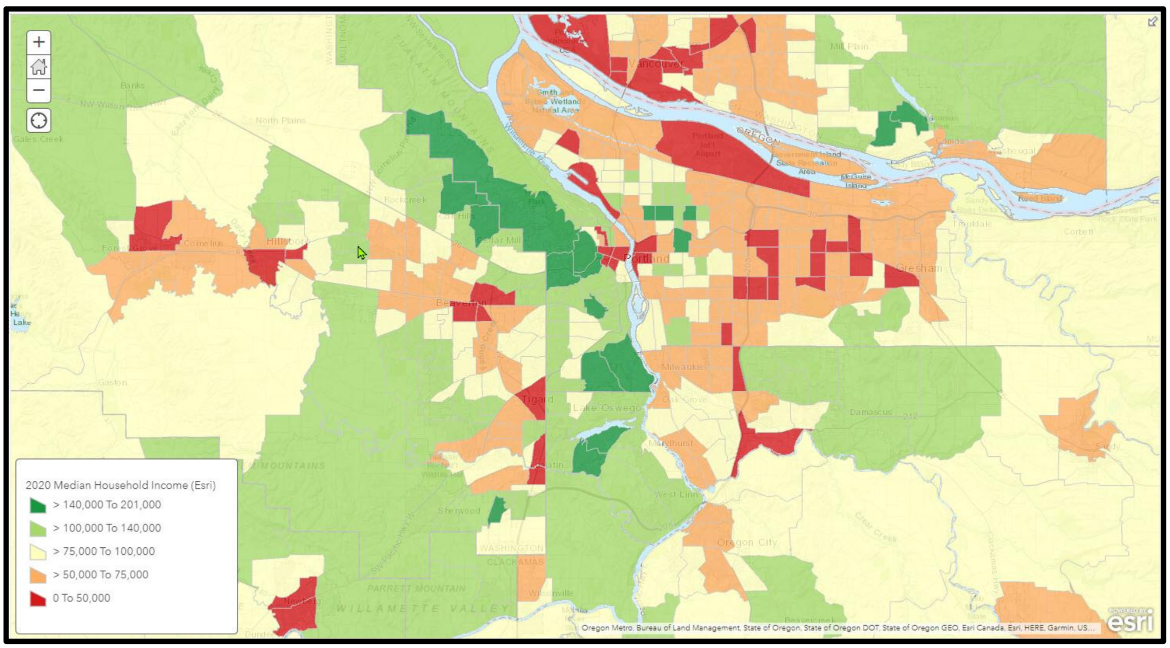

Figure 5 - Median Income: Map of Portland Metro Area 2020 median household income by census tract. (ESRI Community).

Considering Figure 4 again, the green service areas are semi-transparent in order to see areas they cover, which are not covered by red. There are not many, aside from some semi-rural stations. Thus, branded stations appeared to have broad coverage in all parts of the Portland Metro Area, regardless of wealth and income demographics, or at least nearly everywhere that unbranded stations cover. This meant there probably would not be a conclusion that areas of lower wealth have a lack of branded stations, but there might be a conclusion that areas of higher wealth have a lack of unbranded stations.

5.2 Correlation Matrix

Before considering the individual variables, it was important to understand whether the variables chosen had good explanatory power and whether they might be too similar to each other. Since this study dealt with similar indicators – home value, income, net worth – a correlation matrix identifies just how similar these indicators are to each other. Table 3 contains a correlation matrix constructed with each of the eight (8) variables, and the binary "Branding" indicator:

Table 3 - Correlation Matrix

	binary	**medincome**	**dispincome**	**SNAPperc**	**mednetworth**	**medhomevalue**	**wealthindex**	**percapinc**	**avghouseinc**
binary	1								
medincome	0.09	1							
dispincome	0.10	1.00	1						
SNAPperc	-0.04	-0.77	-0.79	1					
mednetworth	0.09	0.83	0.82	-0.51	1				
medhomevalue	0.15	0.70	0.71	-0.56	0.48	1			
wealthindex	0.12	0.93	0.93	-0.70	0.89	0.71	1		
percapinc	0.14	0.85	0.86	-0.68	0.65	0.87	0.84	1	
avghouseinc	0.12	0.97	0.97	-0.75	0.81	0.81	0.96	0.93	1

The correlations are relatively high – which is to be expected, since all indicators are wealth related – with many being in the 70, 80, and even 90 percent range. But there are also a number of less correlated variables. For example, median home value and median net worth only have a 48% correlation, suggesting the value of many people's home is not counted in their net worth, thus they don't have their home fully paid off. Correlations between the binary variable and the indicators are very low, suggesting that by themselves, these variables do not have extremely good explanatory power regarding station branding.

Overall, it would not be a good idea to use these variables in a regular regression because there is high autocorrelation, and there are many variables with important explanatory power that are not included. However, that doesn't mean that conclusive information cannot come from other forms of statistical analysis.

5.3 Results from *t*-Tests

Performing a *t*-test gives a clearer understanding of similarities and differences in the data. For each of the wealth indicators, there were two hundred six (206) branded stations and eighty-two (82) unbranded stations, each with a value that represents its surrounding one (1) mile service area. Because of the type of data, an independent *t*-test with an assumption of unequal variances had to be used. An alpha value of point zero five (.05) was applied. For most of the wealth-based indicators, Income, Wealth, Home Value, etc... the *t*-tests were statistically significant with a one-tailed P-value below point zero five (.05). Only one variable showed an insignificant *t*-test, the percentage of households receiving SNAP benefits. Estimated standard errors of the estimated mean were also calculated for ease of interpretation. Table 4 contains the results from *t*-tests performed on each of the indicators:

Table 4 - t-test results

Median Home Value		
	Branded	*Unbranded*
Mean	410,832	373,361
Est Dev of Est Mean	8,423	10,020
Observations	206	82
df	198	
t Stat	2.88	
P(T<=t) one-tail	0.0022	
t Critical one-tail	1.65	

Wealth Index		
	Branded	*Unbranded*
Mean	97.86	82.35
Est Dev of Est Mean	4.36	3.73
Observations	206	82
df	262	
t Stat	2.71	
P(T<=t) one-tail	0.0036	
t Critical one-tail	1.65	

Median Income		
	Branded	*Unbranded*
Mean	71,992	67,547
Est Dev of Est Mean	1,701	1,843
Observations	206	82
df	217	
t Stat	1.78	
P(T<=t) one-tail	0.0382	
t Critical one-tail	1.65	

Median Disposable Income		
	Branded	*Unbranded*
Mean	55,574	52,439
Est Dev of Est Mean	1,095	1,208
Observations	206	82
df	213	
t Stat	1.93	
P(T<=t) one-tail	0.0274	
t Critical one-tail	1.65	

Per Capita Income

Average HH Income

	Branded	*Unbranded*
Mean	40,118	35,291
Est Dev of Est Mean	1,126	1,275
Observations	206	82
df	208	
t Stat	2.85	
P(T<=t) one-tail	0.0024	
t Critical one-tail	1.65	

	Branded	*Unbranded*
Mean	95,815	87,652
Est Dev of Est Mean	2,305	2,339
Observations	206	82
df	230	
t Stat	2.50	
P(T<=t) one-tail	0.0066	
t Critical one-tail	1.65	

Median Net Worth

	Branded	*Unbranded*
Mean	119,438	84,481
Est Dev of Est Mean	14,178	8,499
Observations	206	82
df	285	
t Stat	2.12	
P(T<=t) one-tail	0.0174	
t Critical one-tail	1.65	

% of Population receiving SNAP

	Branded	*Unbranded*
Mean	0.063	0.066
Est Dev of Est Mean	0.0020	0.0032
Observations	206	82
df	151	
t Stat	(0.70)	
P(T<=t) one-tail	0.2432	
t Critical one-tail	1.65	

There was a highly significant difference in Median Home Value between branded (M=410,832, SE=8,423) and unbranded (M=373,361, SE=10,020) stations; t (198)=2.88, p=.002, suggesting that, on average, the median home value in the vicinity of branded stations is approximately $36,500 higher than that of unbranded stations.

There is also a significant difference between the Wealth Index of branded (M=97.86, SE=4.36) stations and unbranded (M=82.35, SE=3.73); t (262)=2.71, p=.004.

The mean value of 97.86 shows that, on average, the areas surrounding branded stations are almost on par with the rest of the nation, only about two (2) percentage points below. However, those around unbranded are, on average, nearly 18 percentage points below the national average.

Per Capita Income: branded (M=40,118, SE=1,126) and unbranded (M=35,291, SE=1,275); t (208)=2.85, p=.002, and Average Household Income: branded (M=95,815, SE=2,305) and unbranded (M=87,652, SE=2,339); t (230)=2.50, p=.007 are also highly significant. These findings suggest a mean difference in Per Capita Income between branding types of about five thousand dollars ($5,000), and a difference in mean Average Household Income of nearly nine thousand dollars ($9,000). Both of these have a higher mean for branded than unbranded.

Median Income br(M=71,992, SE=1,701) unbr(M=67,547, SE=1,843); t (217)=1.78, p=.038, Median Disposable Income br(M=55,574, SE=1,095) unbr(M=52,439, SE=1,208); t (213)=1.93, p=.027, and Median Net Worth br(M=119,438, SE=14,178) unbr(M=84,481, SE=8,499); t (285)=2.12, p=.017 are also significant at greater than point zero five (.05), but do not have a P-value below point zero one (.01), as the first four variables have. These suggest that Median Income, Median Disposable Income and Median Net Worth have a higher mean surrounding branded stations than unbranded by approximately five thousand ($5,000), three point five thousand ($3,500), and twenty-seven thousand dollars ($27,000), respectively.

While the percentage of households receiving SNAP benefits was not statistically insignificant, with a relatively high P-value around point two five (.25), it still showed a higher mean value for unbranded, as would be expected under the assumption that those with lower income are more likely to receive SNAP benefits, and inverse relationship.

These results provided more evidence that the wealth in service areas surrounding branded stations is higher than in the service areas surrounding unbranded stations.

5.4 Classification and Regression Tree (CART) Models

Although CART models use regression testing at their root, they give different information. One of the conclusions that can be reached about a data set by examining a CART model is which variables have more of an influence on the dependent variable. Those that appear higher up on the tree (closer to the root) are the most influential. Those that occur after multiple splits in the data are still influential, but are less so. Figure 6 depicts a CART model with all eight (8) variables included, and no restrictions or parameters set:

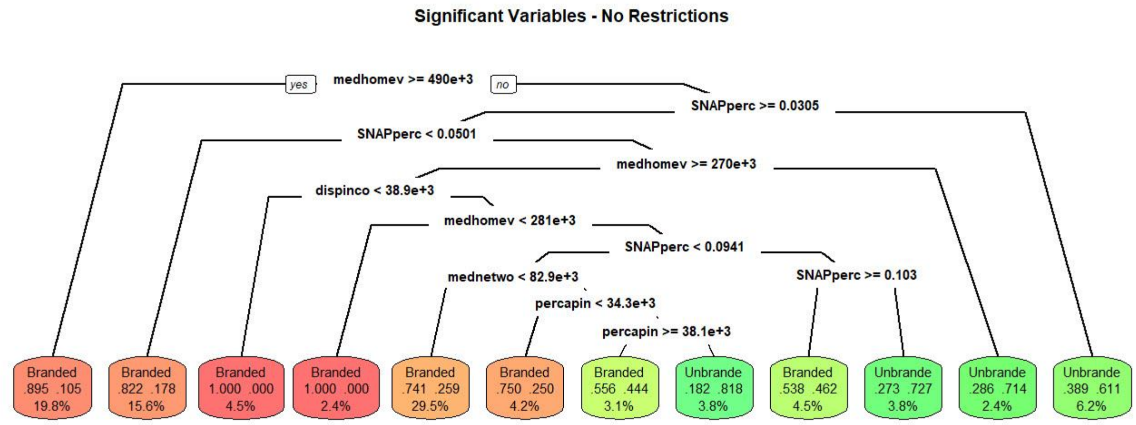

Figure 6 - Unrestricted CART model: Shown with all eight (8) variables included and no restrictive parameters.

Unfortunately, this CART model was a little more complex than what was demanded for this study, and leaving the data in this state actually would have muddled the results due to the fragmentation of the observations, so some restrictions were put in place. First, the variable that was not statistically significant was removed. This resulted in an even more complex tree, so a "max depth" of 6 branches was applied. Figure 7 shows the classification tree with a simple limitation of 6 branches

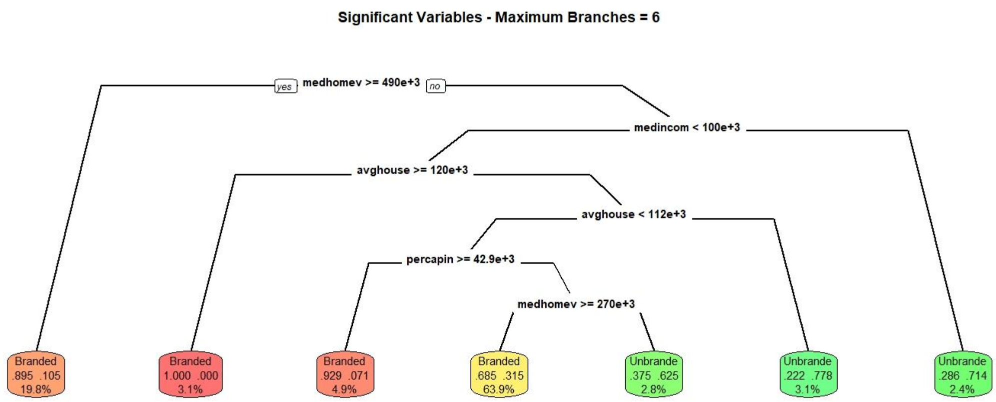

Figure 7 - Restricted CART model: Max Branches = 6. Shown with only statistically significant variables and a maximum depth of 6 branches.

This narrows the field, but is still a lot of information to take in. To further improve the model, a "cost parameter" was set (cp=0.015) so that only more influential splits were made and those that cost the regression in inefficiency were eliminated. The resulting classification tree in Figure 8 is a lot more readable, but is actually exactly the same as the first four (4) branches of Figure 7:

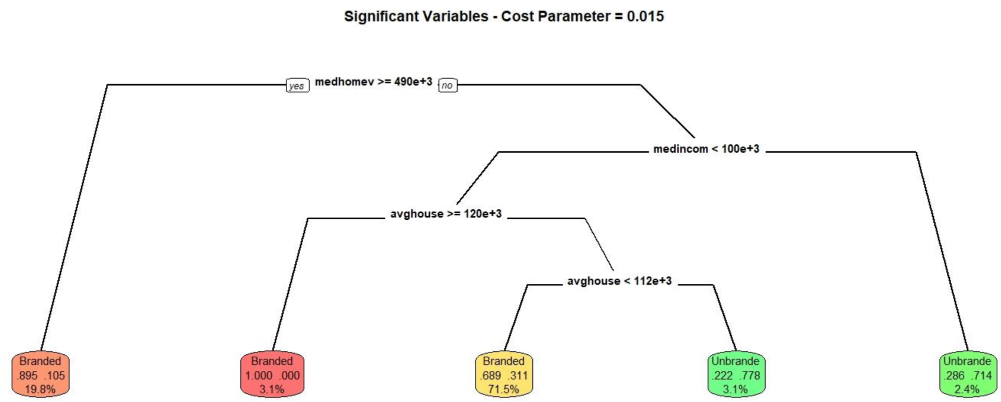

Figure 8 - Restricted CART model: cp = 0.015. Shown with only statistically significant variables and cost parameter (cp) set to 0.015.

CART models are read from the top down. The first branch in the data is with Median Home Values above four-hundred-ninety thousand dollars ($490,000). About twenty percent (20%) of the total number of service areas fall into this category. Among them, approximately ninety percent (90%) are branded. To put it another way, only ten percent (10%) of the stations that are located in areas with a Median Home Value over half a million dollars, are unbranded. This is a very clear statement about the branding choices of stations in areas with high home values: they are mostly branded. No matter how the parameters of the software are adjusted, this variable at this level produces a nearly consistent split with similar figures. This means that it is a very strong branch in the data.

The next most influential variable seen in Figure 8 is Median Income at the one hundred thousand-dollar ($100,000) level. Among service areas with a Median Home Value below four hundred ninety thousand dollars ($490,000), those with an average median income of over one hundred thousand dollars ($100,000) have a propensity score of 71.4% for being unbranded. To qualify this statement, the group only consists of two point four percent (2.4%) of the total number of stations, about seven (7) stations. Of those, approximately 70% are unbranded, about five (5). Since CART diagrams don't give an indication as to which stations are captured in each leaf node, they must be checked manually. Reviewing the map of locations against the data set, these stations appear to be in semi-rural areas where there are slightly higher incomes but only moderately high home values. These stations are not in high population areas.

The third most influential variable is Average Household Income, which splits at one hundred twenty thousand dollars ($120,000). Household Income is a little different measure of income because it includes all incomes for a given household, while Median Income measures each individual income earner. The result of this branch is to separate out three point one percent (3.1%) of the service areas (about nine (9) stations). These service areas are characterized by home values below four hundred ninety thousand dollars ($490,000), Median Incomes below one hundred thousand dollars ($100,000), but an Average Household Income of over one hundred twenty thousand dollars ($120,000). These service areas are one hundred percent (100%) branded stations.

The final branch in Figure 8 is Average Household Income again, but at the one hundred twelve thousand dollar ($112,000) level. Three point one percent (3.1%) of the stations (again, about nine (9) stations), those with greater than one hundred twelve thousand dollars ($112,000) in Average Household Income (but lower than one hundred twenty thousand dollars ($120,000)), have an approximately seventy eight percent (78%) chance of being unbranded.

On the other side of this split is seventy-one point five percent (71.5%) of the total station count, about two hundred six (206) stations. These stations have a sixty nine percent (69%) likelihood of being unbranded. To reiterate, seventy-one point five percent (71.5%) of the stations are located in areas that have an Average Household Income of lower than one hundred twelve thousand dollars ($112,000), and a Median Home Value below four hundred ninety thousand dollars ($490,000). Sixty nine percent (69%) of these stations are unbranded. While these Income and Home values may seem a little on the higher end, these findings still point to the idea that middle and lower wealth areas have more unbranded stations than higher wealth areas.

Another way of using CART models to analyze data is to view each variable individually. This shows where there are natural splits or groupings in the data. For variables that are well distributed and do not have natural break points, this results in no classification tree branches and simply one "root node" instead of a series of "leaf nodes". This is the case for Per Capita Income, Median Net Worth, and Total Population. Other variables may have a number of branches, depending on the distribution of the data points. For this study, all classification trees had an unrestricted cost parameter, but were pruned to three (3), four (4), or five (5) branches, depending on the data. (Five (5) branches for some data may result in a lot more splits than for others. Likewise, only three (3) splits may result in a root node with no branches.) Below are the individual variable models that resulted:

Median Home Value: Branches at four hundred ninety thousand dollars ($490,000), exactly the same as the first branch of the comprehensive model, and again at two hundred seventy thousand dollars ($270,000). The majority of observations are between these two values and have a sixty nine percent (69%) chance of being unbranded.

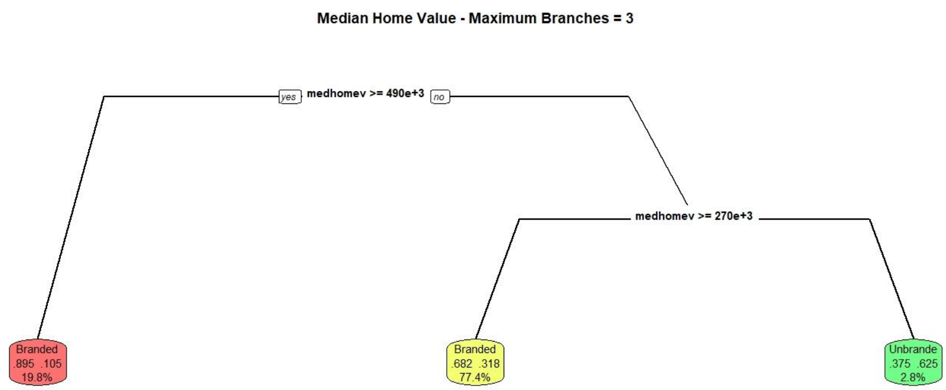

Figure 9 - Median Home Value CART Model: Individual classification tree depicting only Median Income, restricted to maximum of three (3) branches.

Wealth Index: Not wanting to branch less than five (5) times, Wealth Index has a reasonable distribution of values, but still has some breakpoints at very high levels, over one hundred sixty-three (163), and at very low levels, in the ranges of forty-five (45) to eighty (80). The largest segment of observations, nearly forty percent (40%), are between eighty (80) and one hundred sixty-three (163) and are more likely to be branded stations than unbranded.

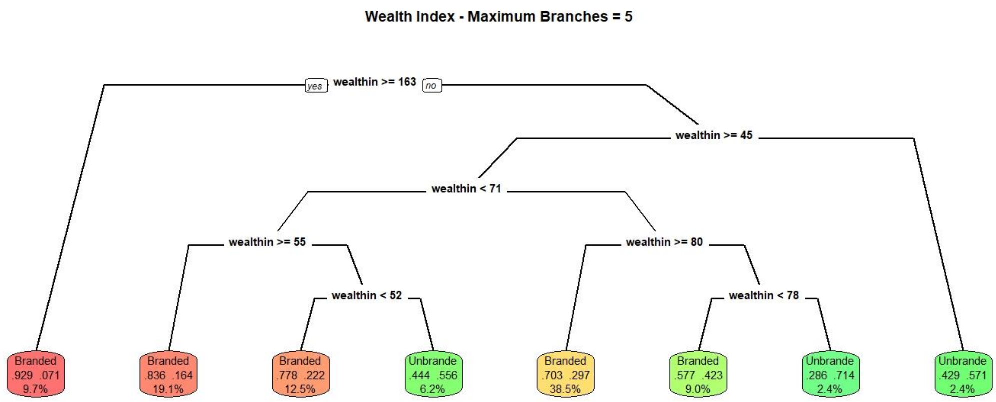

Figure 10 - Wealth Index CART Model: Individual classification tree depicting only Wealth Index, restricted to maximum of five (5) branches.

Average Household Income: Again, not wanting to branch fewer than four (4) times, Average Household Income has an influential split at greater than one hundred thirty-eight thousand dollars ($138,000). All service areas with Average Household Income greater than this value are associated with branded stations, a fact lending itself to the findings that wealthy areas are less likely to see unbranded stations. The remaining leaf nodes show that there is not a lot of clearly differentiated branding based on income.

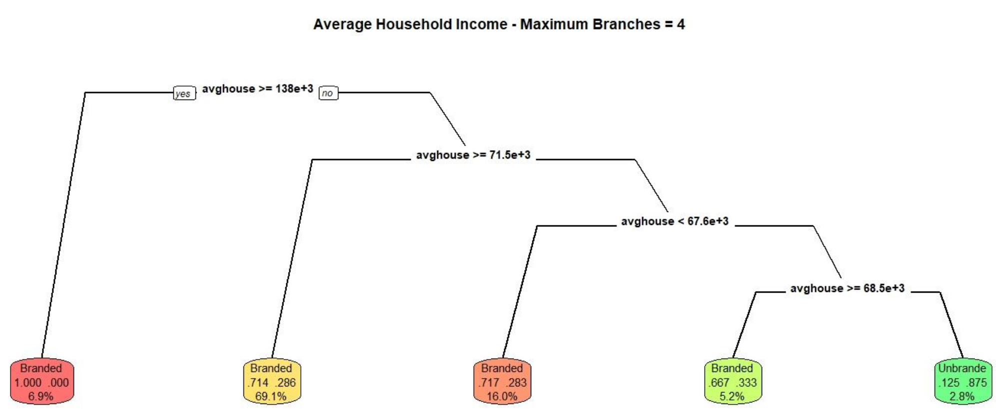

Figure 11 - Average Household Income CART Model: Individual classification tree depicting only Average Household Income, restricted to maximum of four (4) branches.

Median Income: With a strong branch at nearly ninety-two thousand dollars ($92,000), branded stations make up nearly eighty-five percent (85%) of the forty-seven (47) stations which are above that threshold. The remaining grouping are less differentiated and of less consequence.

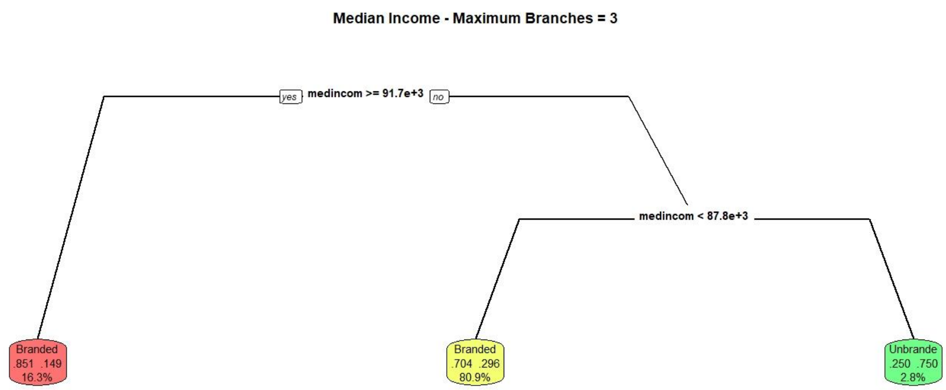

Figure 12 - Median Income CART Model: Individual classification tree depicting only Median Income, restricted to maximum of three (3) branches.

Disposable Income: Also, not wanting to branch less than four (4) times, Disposable Income has a high end split similar to Median Income, where stations above the threshold of about sixty-eight thousand dollars ($68,000), forty-seven (47) of them, have a nearly eighty-five percent (85%) likelihood of being branded. Similarly, the remaining groupings are varied enough that their figures are of less consequence.

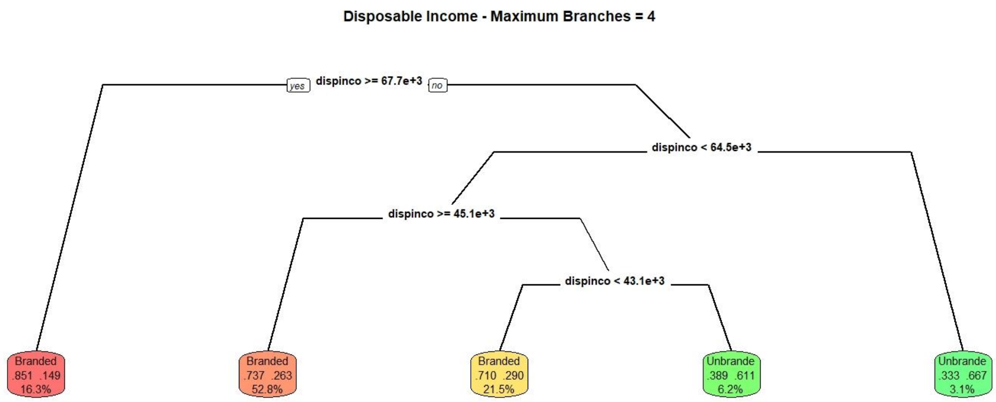

Figure 13 - Disposable Income CART Model: Individual classification tree depicting only Disposable Income, restricted to maximum of four (4) branches.

The remaining variable was not statistically significant, so its classification tree was not included.

When comparing each of these analyses, the general conclusion supports the idea that lower wealth regions have a higher predominance of unbranded stations than higher wealth areas; and that areas of higher wealth have a distinct lack of unbranded stations, while most other areas have a good representation of both branded and unbranded stations.

Section 6: Discussion and Conclusions

The research question presented herein was whether the wealth demographics of branded and unbranded gas station service areas are different. The initial expectation was that areas surrounding unbranded stations would be found to have generally lower wealth characteristics than branded. The preponderance of the results from these analyses point to just such a conclusion.

The first analysis, a simple visual comparison of service area maps with branded stations layered on top of unbranded against a map depicting median income, showed that branded stations are well represented in all neighborhoods in the Portland Metro Area, particularly those in which unbranded stations also exist. However, unbranded stations are not well represented in all areas. In particular, the regions colloquially known as the "wealthier parts of town", and showing higher median income, seem to be nearly devoid of unbranded stations.

The second analysis, a series of two-tailed *t*-tests, showed a statistically significant difference between a variety of wealth demographics in the service areas of branded and unbranded stations. One-tailed tests affirmed not only a significant difference, but that the indicators: median home value, wealth index, per capita income, average household income, median income, median disposable income, and median net worth, all showed branded service areas having a statistically significant higher mean value than that of unbranded. While a correlation matrix showed that some of these indicators give similar results, due to their descriptive similarity, each was different enough that it lends further legitimacy to the results.

The third analysis, a CART model, in particular a classification tree, told a story about the natural splits in the data and how influential each indicator might be in a regression model. Home value was found to be the most influential indicator, with high home value regions unlikely to see unbranded stations, compared to lower home value areas. Moderate to lower value homes with moderately lower incomes were found to be more likely to see unbranded stations, which also supports this conclusion. Individual variable classification trees also support the conclusion that wealthier areas are far less likely to see unbranded stations.

These findings lead to the confirmation of the economic assumption that area wealth and station branding do have some correlation. They do not necessarily support the conclusion that unbranded stations are located exclusively in lower wealth areas, and branded stations are not, as branded stations are also present in lower wealth areas. But it does support the notion that higher wealth areas, particularly very high wealth areas, are unlikely to host an unbranded station.

This means that lower wealth individuals and families may not necessarily be subjected to price volatility and a potential for shortages due to their use of unbranded stations, since they may well have branded options nearby. Meanwhile, higher wealth individuals and families may not face this speculative volatility and these product shortages. However, the prices they face are not necessarily set by competitive practices.

In trying to link these findings with other geospatial concepts of urban analysis, Central Place Theory (CPT) would see the distribution of branded stations as entirely normal: more centrally-located people would demand more centrally-located commodities, and a high availability of goods and services would tend to draw in more customers and residents. When examining the relationship between urban and rural station locations, without considering branding, this would seem to hold true.

The lack of unbranded stations in higher wealth areas might defy this concept, but the presence of gasoline in any region has downsides. It is smelly, toxic and flammable. It is culturally understood by many as a necessary evil, something that without doubt enhances lives, but is also something to keep at arm's length due to its potentially negative side-effects. An avenue for further research which might take from, and add to, concepts of socio-spatial research, could be to analyze how gasoline branding might cause NIMBY, (Not In My Back Yard), responses to either newly proposed fueling stations, or to the continued operation of what may be perceived as outdated and unsafe facilities. Since new gas stations are rarely built within the Portland metropolitan area, gentrification and NIMBY-ism would seem likely to play a role in shaping the location and branding of gas stations over time.

Introducing time into the analysis adds multiple layers of additional data collection and processing. While this study was meant as a snapshot in time, (an observational analysis of the conditions that currently exist in the Portland Metro Area), this same data set, sampled at a variety of times over the past decade or two would result in a collection of "time-series" panels. A comprehensive time-series of gas station location data may reveal whether there is a connection between a rising level of localized wealth and the occurrence of unbranded station owners refurbishing their stations to obtain a branding contract; or in more extreme cases, station owners shutting down refueling operations in favor of a café or mini-mart. It may even be possible to pinpoint certain NIMBY actions taken by locals who were publicly vocal about their desires.

To recap, this study shows that wealth indicators do not seem to affect branded stations, as they tend to have good coverage in all parts of the Portland metro area. However, unbranded stations do seem to be affected, in that they are generally not located in higher wealth areas. It is unclear why wealthier areas seem to avoid unbranded stations, but with a time-series of data, some of the reasons may be revealed.

Thinking toward generalizability and whether it is possible to conduct this study in other metropolitan areas, it is unclear if things like geography, local land use regulations, use of urban growth boundaries, or even simple social differences would affect the outcome. Portland has a somewhat centralized population with relatively sparse populations in immediately adjacent rural areas, and an urban growth boundary that forces local land use to be carefully considered. If instead, a rapidly expanding city were to be analyzed; one that doesn't have urban growth boundaries, one that is interested in building new refueling stations in suburbs, exurbs, and satellite cities, one with a very different distribution of wealth across their area, the results might be very different. Conversely, a study of an older and more geographically isolated city, that is even more restrictive in their land use laws than Portland, has very little room for expansion or need of new gas stations, and has fairly segregated communities based on home value or income, might show an even stronger correlation between branding and wealth.

Regardless of those potential outcomes, one thing is clear. These characteristics of demand, location, and branding changes are well within the scope of a properly functioning free market. Consumers make their demands known, and business owners pivot to accommodate those demands. It is not the intention of this analysis to make normative statements about where unbranded stations “should” locate, nor about any corrective measures that city planners “should” take. This paper was merely intended as an observational study using a newly generated data set and method of analysis.

Section 7: References

AAA. (2016). *AAA Fuel Quality Research Proprietary research into the effectiveness of fuel additive packages in commercially-available gasoline* (Rep.). Retrieved 2020, from American Automobile Association website: https://www.aaa.com/AAA/common/AAR/files/Fuel-Quality-Full-Report.pdf

ABC News. (2005, October 29). What's the Difference Between Regular and Discount Gas? Retrieved 2020, from https://abcnews.go.com/GMA/Consumer/story?id=1261790

Bergeaud, A., & Raimbault, J. (2020). An empirical analysis of the spatial variability of fuel prices in the United States. *Transportation Research Part A: Policy and Practice, 132*, 131-143. doi:10.1016/j.tra.2019.10.016

Clover, V. T. (1953). Price Influence of Unbranded Gasoline. *Journal of Marketing, 17*(4), 388. doi:10.2307/1247015

Doyle, J. J., & Samphantharak, K. (2008). $2.00 Gas! Studying the effects of a gas tax moratorium. *Journal of Public Economics, 92*(3-4), 869-884. doi:10.1016/j.jpubeco.2007.05.011

EIA. (2015, September). PADD 5 Transportation Fuels Markets. Retrieved 220, from https://www.eia.gov/analysis/transportationfuels/padd5/pdf/transportation_fuels.pdf

ESRI. (2020). Comparison Reports, Community Analyst. Retrieved 2020 from https://doc.arcgis.com/en/community-analyst/help/comparison-reports.htm

ESRI. (2020). Generate Service Areas (Ready To Use). Retrieved 2020 from https://pro.arcgis.com/en/pro-app/tool-reference/ready-to-use/itemdesc-generate-service-areas.htm

ESRI. (2020). Exploring Esri's Wealth Index. Retrieved 2020, from https://storymaps.arcgis.com/stories/f159a3525b964105990872ea17d4f084

Exxon Mobil. (2020). The Journey of Fuel. Retrieved 2020, from https://www.exxon.com/en/fuel-journey

Google Maps. (2020). Street View. Retrieved 2020, from https://support.google.com/maps/answer/3093484?co=GENIE.Platform=Desktop&hl=en

Kendrick Oil. (2017, June 02). How Branded Fuel Stations Operate. Retrieved 2020, from http://www.kendrickoil.com/how-branded-fuel-stations-operate/

Kendrick Oil. (2017, July 22). The Differences Between Branded Vs. Unbranded Fuel - Kendrick Oil Company. Retrieved 2020, from https://www.kendrickoil.com/branding/the-differences-between-branded-vs-unbranded-fuel/

Kendrick Oil. (2017, July 22). Price Inversions: What Are They And Why Do They Happen? - Kendrick Oil Company. Retrieved 2020, from https://www.kendrickoil.com/products-services/price-inversions-what-are-they-and-why-do-they-happen/

Marathon. (2019). Annual 10K Report: Marathon Petroleum Investors. Retrieved 2020, from https://www.marathonpetroleum.com/Investors/Annual-Report-Proxy-Statement/

National Association of Convenience Stores (NACS). (2020, February 17). How Branded Stations Operate. Retrieved 2020, from https://www.convenience.org/Topics/Fuels/How-Branded-Gasoline-Stations-Work

OPIS & Berhang, S. (2017). Pricing 101 Part 1: Your Basic Guide to Pricing Gasoline and Diesel. Retrieved 2020 from http://blog.opisnet.com/pricing-101-your-basic-guide-to-pricing-gasoline-and-diesel

OPIS, & Berhang, S. (2017). Pricing 101 Part 2: Spot Fuel Markets Made Simple. Retrieved 2020, from http://blog.opisnet.com/spot-fuel-markets-made-simple

OPIS, & Berhang, S. (2017). Pricing 101 Part 3: Wholesale Rack Fuel Pricing Essentials. Retrieved 2020, from http://blog.opisnet.com/wholesale-rack-fuel-pricing-essentials

OPIS, & Berhang, S. (2017). Pricing 101 Part 4: Demystifying Retail Fuel Prices and Players. Retrieved 2020, from http://blog.opisnet.com/demystifying-retail-fuel-prices-and-players

ReferenceUSA (formerly). (2020, October 27). Data Axle Reference Solutions. Retrieved 2020, from https://www.data-axle.com/what-we-do/reference-solutions/

Spatial Data Library. (2019). Oregon Transportation Network 2019. Retrieved 2020, from https://spatialdata.oregonexplorer.info/geoportal/details;id=161c9568b5584a6d9ff9b019da09f034

Spatial Data Library. (2014). Metro Boundaries. Retrieved 2020, from https://spatialdata.oregonexplorer.info/geoportal/details;id=6fb5d387f8e94ce5917f384c82964219

Tesoro. (2006). *2005 Annual Report* (pp. 1-121, Rep.). San Antonio, TX: TSCorp. Retrieved 2020, from https://www.sec.gov/Archives/edgar/vprr/0603/06033897.pdf

Yse, D. L. (2019, April 30). The Complete Guide to Decision Trees. Retrieved 2020, from https://towardsdatascience.com/the-complete-guide-to-decision-trees-28a4e3c7be14