



INDIANA UNIVERSITY
**SCHOOL OF
INFORMATICS, COMPUTING,
AND ENGINEERING**

Project B: Big Data Concepts

Author:
Raymond Tan

TA:
Utkarsh Kumar

Instructor:
Dr. Inna Kouper

A Project submitted for:

I-535 Management, Access, and Use of Big and Complex Data

February 12, 2020

1. INTRODUCTION

This Data Analytics project is from real data with some information anonymized for confidentiality. An online car dealership in Atlanta GA has requested our company to provide some insights to their business.

CarTrader wants to know how they are performing with their listings in the Atlanta market compared to the competition. They are most concerned about the used car marketplace because they have not seen a lot of data in that area.

2. BACKGROUND

This project is really interesting because it is a real business problem and will give insights to used car market in Atlanta Georgia. We will analyze the issue with the online dealership, the competition and provide some recommendation for CarTrader.

3. METHODOLOGY

To get the performance of CarTrader listings, we we pulled 3 database tables worth of 1 year data from October 2018 to Oct 2019: These tables comes from one of the biggest company that tracks vehicle transactions.

1. Customer Lookup Sample File.txt – This file contains the information about the customer that will be used for comparison
2. Vehicle Lookup Sample File.txt – This file contains the information about the vehicles
3. Vehicle Details Sample File.txt – This is the transaction log file that contains user activity for new and used vehicles on our Website

The tables were processed, inspected, and analyzed for data consistency in OpenRefine. The data was then further processed in Tableau, taking the customer data as a primary table. We left-join the Details table using customer ID. To get the vehicle information, we also left join the Vehicle lookup table to the combined table using Vehicle ID.

Figure 1: OpenRefine cleaning

There are 2 important parameters for this study. The analysis is targeted for the used car market in Atlanta. We have to ensure that we have filtered the data with used car information only. Another parameter is that the analysis is for the Atlanta market only. The available data are all for the Greater Atlanta market, so no further filtering of data was required.

4.1. Price to Market Index

CarTrader is pricing its used cars well. However, more Sites are pricing their cars more aggressively. We ranked 6th, which means that people will tend to go to other sites when they are looking for better deals in used cars. We could further decide to be even more aggressive in our pricing if we want to compete with PMI.

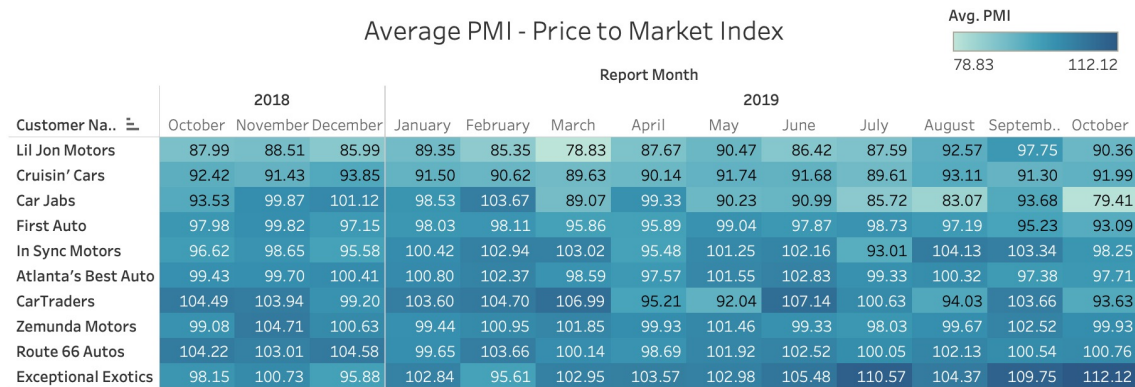


Figure 2: Average PMI

4.2. Scarcity

Scarcity – This is an indicator of how much supply of a vehicle there is compared to the demand for that vehicle. A higher value would indicate that there is more demand than supply. A lower number would indicate more supply than demand.

We are ranked 3rd in scarcity. Our scarcity numbers mean that there's a lot of demand in used cars, and CarTrader has a high shortage of used cars. We have to find a way to put more vehicles in our used car inventory.

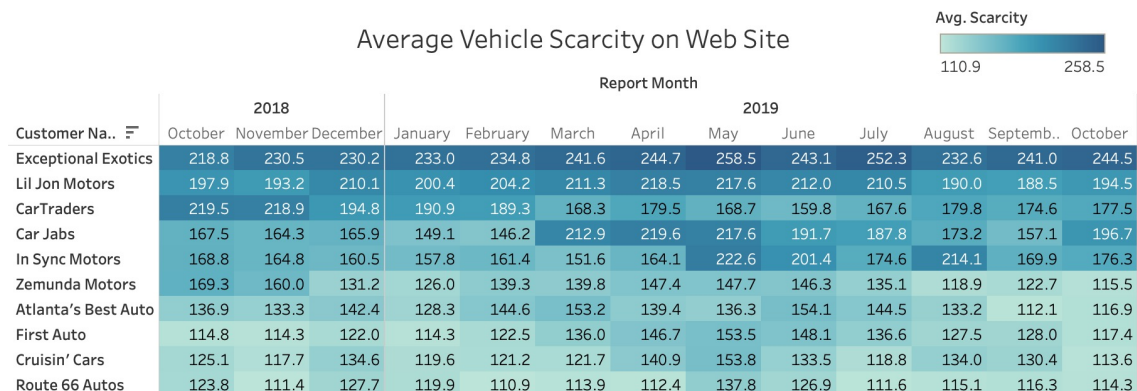


Figure 3: Average Scarcity of Vehicles in Website

4.3. Days on Site

Days on Site - This is an indicator of how long in the number of days a vehicle is available from the Website. Smaller number will indicate that we are selling cars faster. This number could also mean that we are just quicker updating our Website when the vehicles are removed from our website.

For Days on Site, we are ranked 8th. This could indicate that we are not selling as fast as our competitors. This could also mean that we are not updating our website quickly soon enough when we sell a used car.

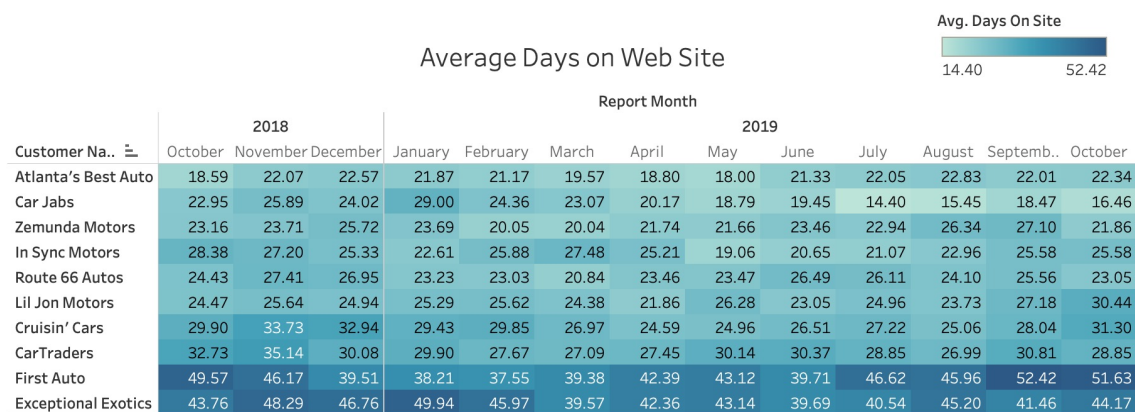


Figure 4: Average Days on Site

4.4. Search

Search - This is an indicator of how many people are searching for our website for a used car. More search means our site is more popular for people who are looking to buy a used car. A low number means our website is not as popular as other sites.

For search, we are ranked last. This could indicate that our website is the least popular for people when they are looking to buy used car. This could signal that our website is not user-friendly, not marketed, we do not have inventory of what people wants, our price is higher or social media reviews may not be positive.

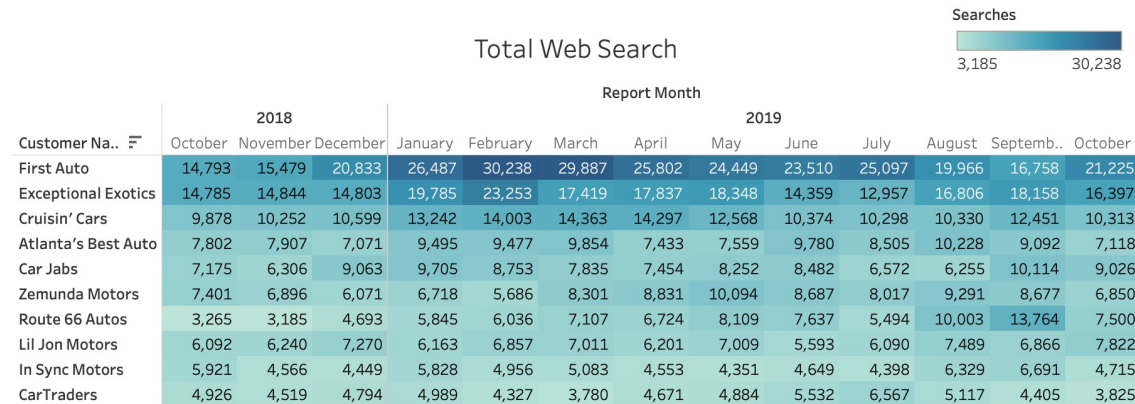


Figure 5: Total Search made in Website

4.5. Most Searched Vehicles

We need find out what people are searching most. CarTraders is last on search. This will give us a gauge of the demand of used car market, which type and brand are most searched. It was fascinating to find that around 80% of the searches are all year models of Ford Trucks and Jeep SUV's.

Most Searched Vehicles

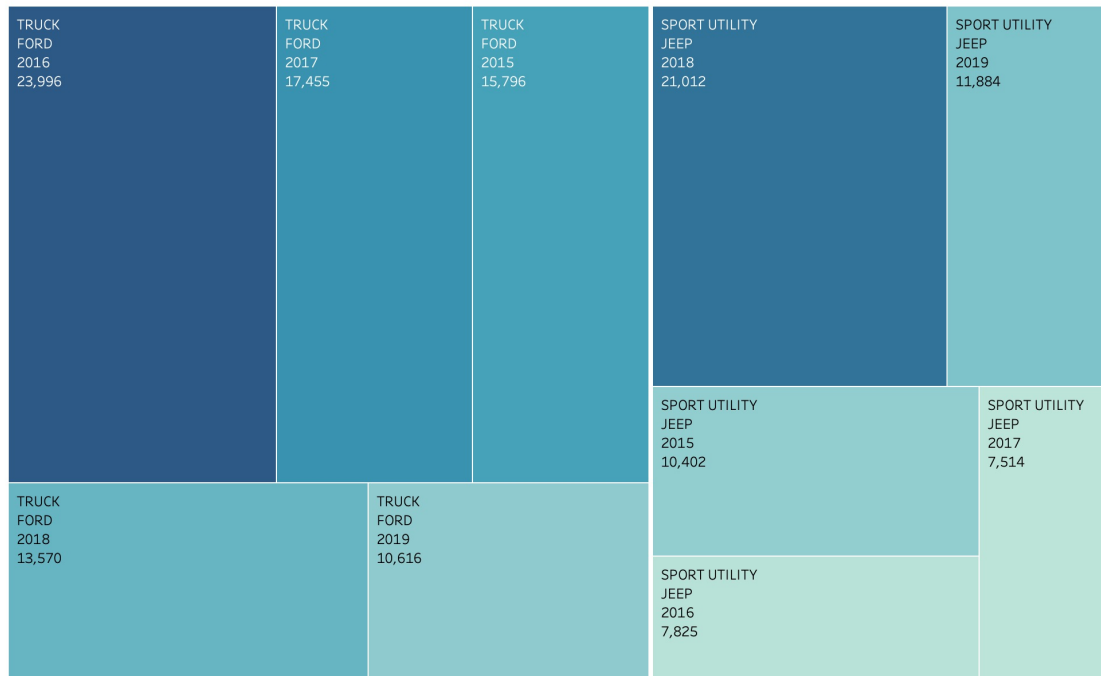


Figure 6: Over 80% of Search was Jeep SUV and Ford Trucks

4.6. Inventory Analysis

People are searching for used Ford Trucks and Jeep SUV's. In this inventory analysis, we analyzed the inventory of the user cars market. CarTraders are lagging behind in this 2 car segments which we think contributes to the lack of searches and popularity with people looking for used car in the internet.

Inventory for Jeep SUVs

Customer Name	Report Month												
	2018			2019									
	October	November	December	January	February	March	April	May	June	July	August	September	October
Customer Name	31	30	31	31	28	31	30	31	30	31	31	30	31
Cruisin' Cars	339	245	186	175	178	203	240	287	415	575	569	556	400
Exceptional Exotics	355	286	246	196	177	187	221	221	241	340	335	330	217
Route 66 Autos	232	169	145	127	120	147	135	142	164	273	270	247	159
Lil Jon Motors	223	186	162	144	142	157	154	145	163	233	204	207	151
Zemunda Motors	3		1				1	3	7	13	14	14	9
First Auto	15	11	7	1	1	2	2	2	3	5	7	4	3
Atlanta's Best Auto	5	1							2	14	11	10	8
Car Jabs	5	2							2	7	4	9	6
In Sync Motors	4	2							1	3	3	6	3
CarTraders								2	3	4	4	3	1

Figure 7: Inventory Jeep

Inventory for Ford Trucks

Customer Na. 𐄂	Report Month												
	2018			2019									
	October	Novem..	Decemb..	January	February	March	April	May	June	July	August	Septem..	October
	31	30	31	31	28	31	30	31	30	31	31	30	31
Lil Jon Motors	512	355	303	283	298	336	428	542	701	939	1,059	1,110	864
First Auto	217	146	117	100	94	95	128	182	268	357	409	395	285
Car Jabs	212	134	112	96	105	134	173	187	205	261	279	294	176
In Sync Motors	140	89	75	76	72	81	94	108	120	163	176	163	126
Cruisin' Cars	6	2	2					1	6	12	14	13	6
Zemunda Motors	6	3	1	1				2	5	14	14	8	1
Exceptional Exotics	3	1	1					1	4	11	14	10	6
Atlanta's Best Auto	4	1							5	10	10	11	8
Route 66 Autos	2									1	4	5	1
CarTraders											2	2	1

Figure 8: Inventory Ford

CarTraders almost do not have the most popular used cars. What does the type of cars CarTraders has and their competitors? From our analysis, CarTraders have most inventory on sedans, convertibles and wagons. CarTraders need to have more Ford Trucks and Jeep SUV to be able to compete.

Inventory By Type

Customer Name	Type							
	CONVE..	COUPE	HATCH..	SEDAN	SPORT UTILITY	TRUCK	VAN	WAGON
Atlanta's Best Auto	57	766	248	1,332	2,034	3,195	363	9
Car Jabs	46	378	47	412	1,717	2,477	185	6
CarTraders	287	252	9	1,516	1,722	10	345	81
Cruisin' Cars	33	404	33	744	5,214	3,142	1,126	15
Exceptional Exotics	100	236	47	217	3,797	1,360	311	14
First Auto	137	514	186	1,033	3,031	3,032	915	22
In Sync Motors	72	724	56	286	1,031	1,535	153	
Lil Jon Motors	63	766	166	1,331	6,842	11,107	1,566	7
Route 66 Autos	16	304	18	583	2,916	1,167	208	15
Zemunda Motors	67	571	138	711	2,050	2,741	289	2

Figure 9: Inventory

Ranking by Inventory Type

Customer Name	Type							
	CONVE..	COUPE	HATCH..	SEDAN	SPORT UTILITY	TRUCK	VAN	WAGON
Atlanta's Best Auto	7	1	1	2	7	2	4	6
Car Jabs	8	7	6	8	9	6	9	8
CarTraders	1	9	10	1	8	10	5	1
Cruisin' Cars	9	6	8	5	2	3	2	3
Exceptional Exotics	3	10	6	10	3	8	6	5
First Auto	2	5	2	4	4	4	3	2
In Sync Motors	4	3	5	9	10	7	10	
Lil Jon Motors	6	1	3	3	1	1	1	7
Route 66 Autos	10	8	9	7	5	9	8	3
Zemunda Motors	5	4	4	6	6	5	7	9

Figure 10: Inventory Ranking

5. DISCUSSION

Looking at the results, the inventory of the CarTraders are not what the people wants. There's also an indication that the pricing and availability could contribute to the poor performance of CarTrader.com compared to competitors.

In this project, I implemented the data lifecycle we discussed in week 4. I initially thought of putting the data in our JetStream virtualized environment and use DFS but the 1 year data is structured and can be handled by Tableau directly. I implemented Quality and Cleaning, Analytics and implemented Goal setting, understanding data and implemented tools to get desired project goals.

If this will be a recurring project or if continuous analysis will be required by the customer, we'll re-design a data pipeline that will pull this 3 database automatically, put them in a RDBMS or MongoDB. We'll pre-process them in Alteryx and probably design an automated analytics product and use Tableau for ad hoc analytics request.

Our vehicle database also has some duplication and I have informed our data engineering team to look at them and clean them.

6. CONCLUSION

We suggested CarTrader to overhaul their inventory and put more Ford Trucks and Jeep SUV used cars in their Website We also suggested if they could be a little more aggressive with pricing specially as they try to rebuild their website traffic. We recommend some form of research in Atlanta used car market because we don't know if there are UX issues with the website as well, could be focus groups, commissioning ethnographers, or UX research. We discussed a need for Thick Data of people's used car buying trends to understand why they are not searching at CarTraders.com. Some of the questions we could ask are: Where will they go if they will look for a used car? What is it that they like in a website when buying a used car?

What is Thick Data?

Thick Data. The importance of context and emotion

The term Thick Data has been popularized by the anthropologist Tricia Wang and it refers to "dense data", a more than evident nod to Clifford Geertz's "dense description". The Thick Data differs from Big Data by its qualitative approach, obtaining ethnographic data that allows to reveal contexts and emotions of the studied subjects. While Big Data requires an algorithmic process usually carried out by statesmen and mathematicians, Thick Data is the ground of anthropologists, sociologists, and social scientists.¹

¹Valero, What is Thick Data?. April 17, 2017, <https://blog.antropologia2-0.com/en/>

“

When organizations want to build stronger ties with stakeholders, they need stories. Stories contain emotions, something that no scrubbed and normalized dataset can ever deliver.

Numbers alone do not respond to the emotions of everyday life: trust, vulnerability, fear, greed, lust, security, love, and intimacy.

It's hard to algorithmically represent the strength of an individual's service/product affiliation and how the meaning of the affiliation changes over time.

Figure 11: Thick Data