BIA 610 - APPLIED ANALYTICS

Prof. VICTOR MOYA

DATA ANALYTICS FOR LAUNDRY BUSINESS EXPANSION

Submitted by -

KAVIT SHAH

RAJ MAHAJAN

SHUBHAM SHRIVASTAVA

---

1. Business Understanding

An existing laundry chain wants to build a vast network in small cities within the US to maximize profit and to avoid competition with big players in major cities. The company is already located in 140 locations and recently opened stores in 10 new cities. They want to identify which of the 10 new locations have the best potential for the company to invest more funds into marketing.

We are addressing the problem using Tableau by implementing clustering, building trend lines and regression to identify the location that would provide the most profit.

Data Understanding

We have used two datasets for the project, which are as follows: -

1. P1-Startup Expansion
2. P1-US Cities Population

The selected datasets consist of 6 dimensions and 6 measures that are used in the evaluation. The data contains mostly textual data like Sales region, City, Money spend on marketing, and revenue made. The US-population data contains data about cities and their respective population. There is no need for data normalization since most of the data is textual and numerical data is very less. Also, there are no black values in the data which make data preprocessing easy.

2. Data Preparation

Calculate fields are generated manually for improved visualizations and to carry out the data analysis.

- Return on Marketing Investment

- We have also removed duplicate copies of the data entry while joining (inner joins) the datasets.

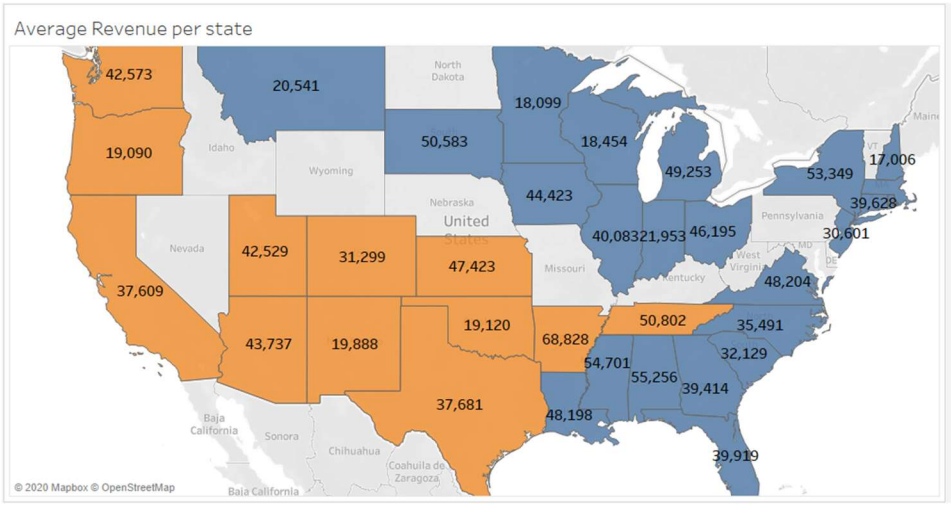
3. Modeling

After performing refined clustering, we found that the yellow line is highly profitable for the company. By looking at the world map with cities mapped with respective clusters, it is easy to point out the cities to choose from that will be highly profitable for the company. Most of the new stores that should interest the company's decision-makers lie in California and the company should plan to invest more in those areas. The model generated would be useful for the company to make a strategy for marketing.

4. Evaluation

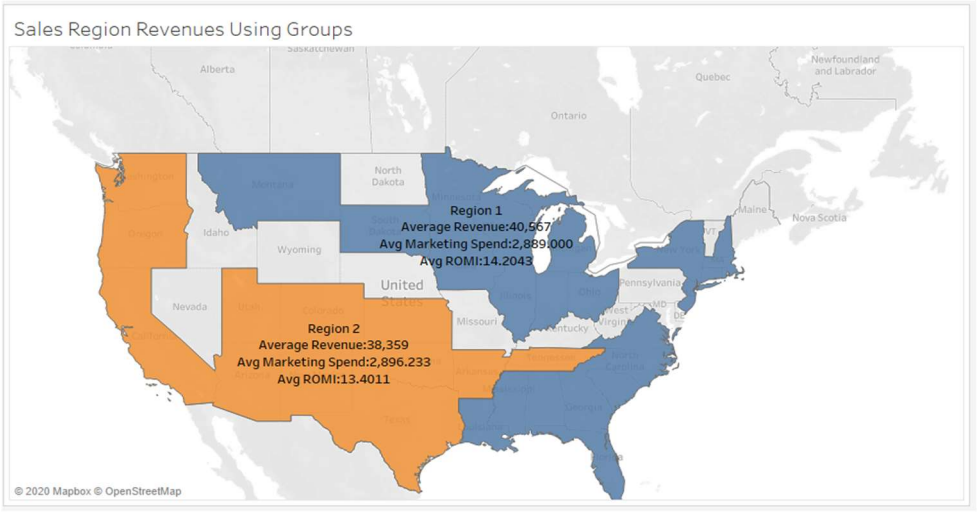
The following screenshots describe the results of our findings: -

1. Average Revenue Per State



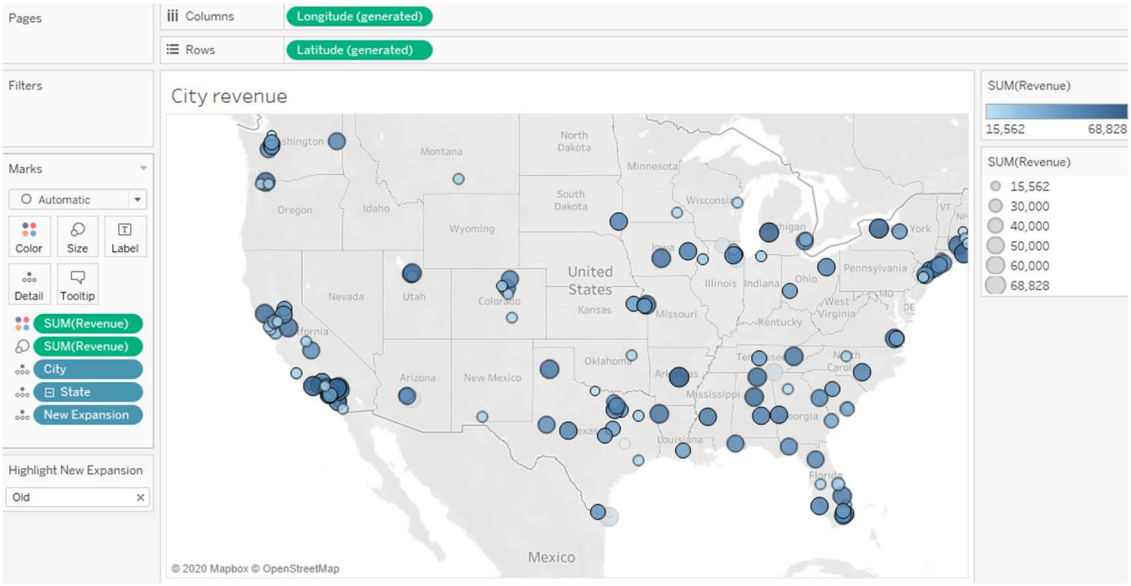
Here we can see the Average Revenue per state of the Laundry Startup in the USA. The Sales region is later divided into colors.

2. Sales Region Revenues Bifurcation



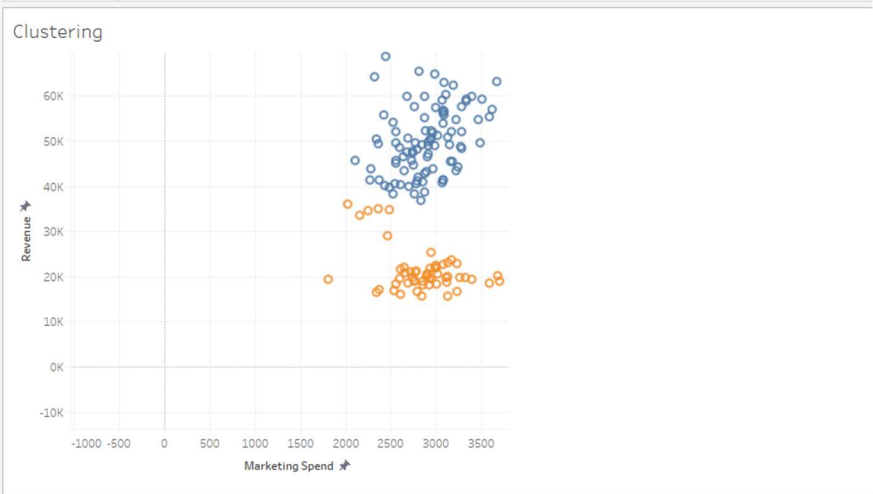
Here the different sales regions' revenue is bifurcated using group function in Tableau to get a better idea of the network of the Laundry Startup.

3. City Revenue



Here we increase our level of Granularity by adding the cities with their contribution to the revenue. We use the high-lighter feature to select from 140 old cities or 10 new cities.

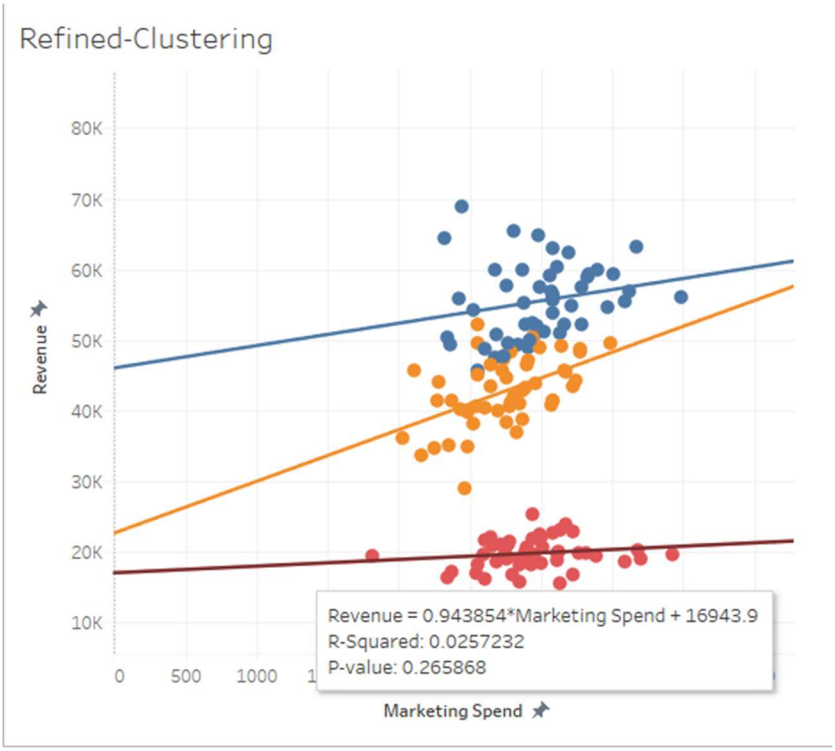
4. Clustering



We have done simple clustering between Revenue and Marketing spend but as the startup like this depends more on the number of people it is necessary for us to include population as a factor too. Therefore we need to refine the cluster further.

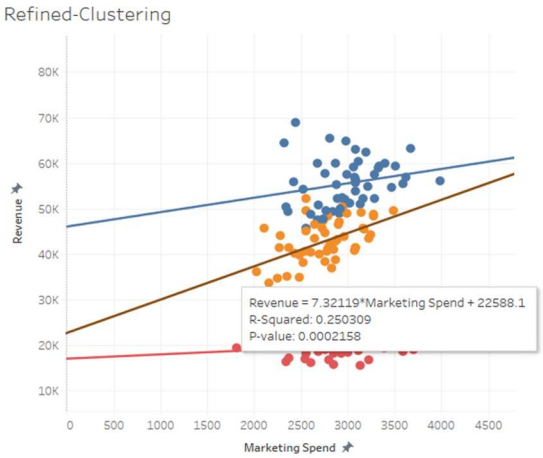
5. Refined Clustering

- **Refined Clustering for Red Trend Line**



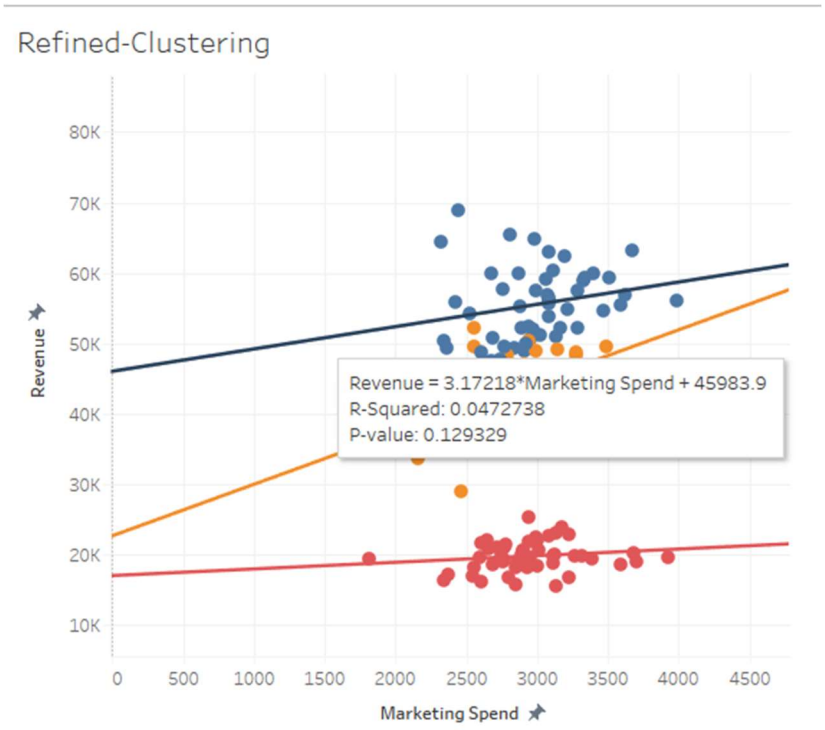
For this refined clustering, we have included population as a factor too and the trend line helps us see different cities with the potential to invest. The red trend line signifies that for every $1 spent on marketing the returns are $0.934 which is not good.

- **Refined Clustering for Yellow Trend Line**



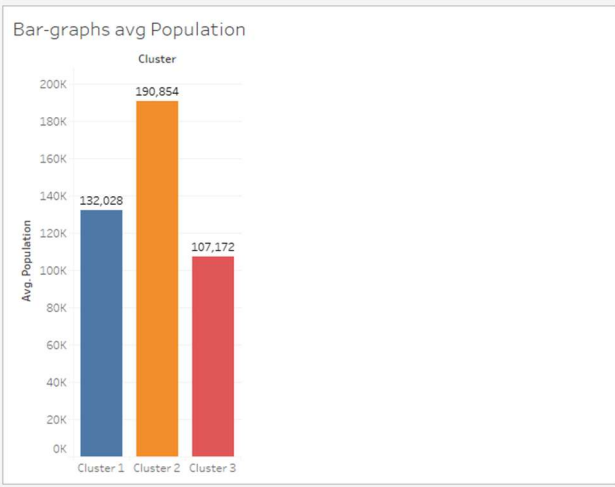
Now for the Yellow Line, it signifies that for every $1 spent on marketing the returns $7.32 which is high and profitable. Compared to the remaining lines it is the best trend line which tells us about the cities which have high potential to invest in marketing funds.

- **Refined Clustering for Blue Trend Line**



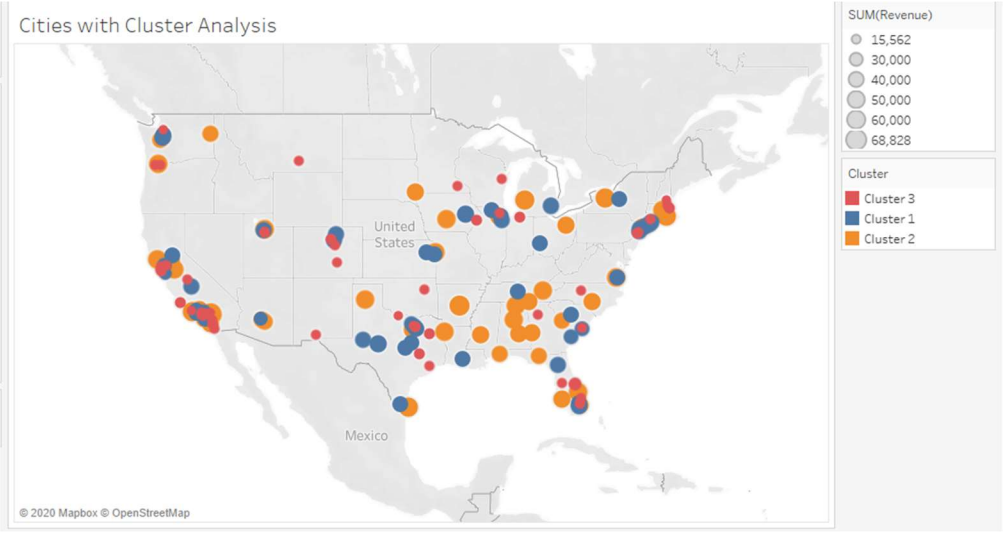
For the blue trend line where the population is the maximum, it signifies that for $1 spent on marketing the returns are $3.17 which is good but not that good enough as the yellow line. So these trend lines help us select the new cities which are available or have the potential to invest the money for marketing.

6. Bar Graphs for Population



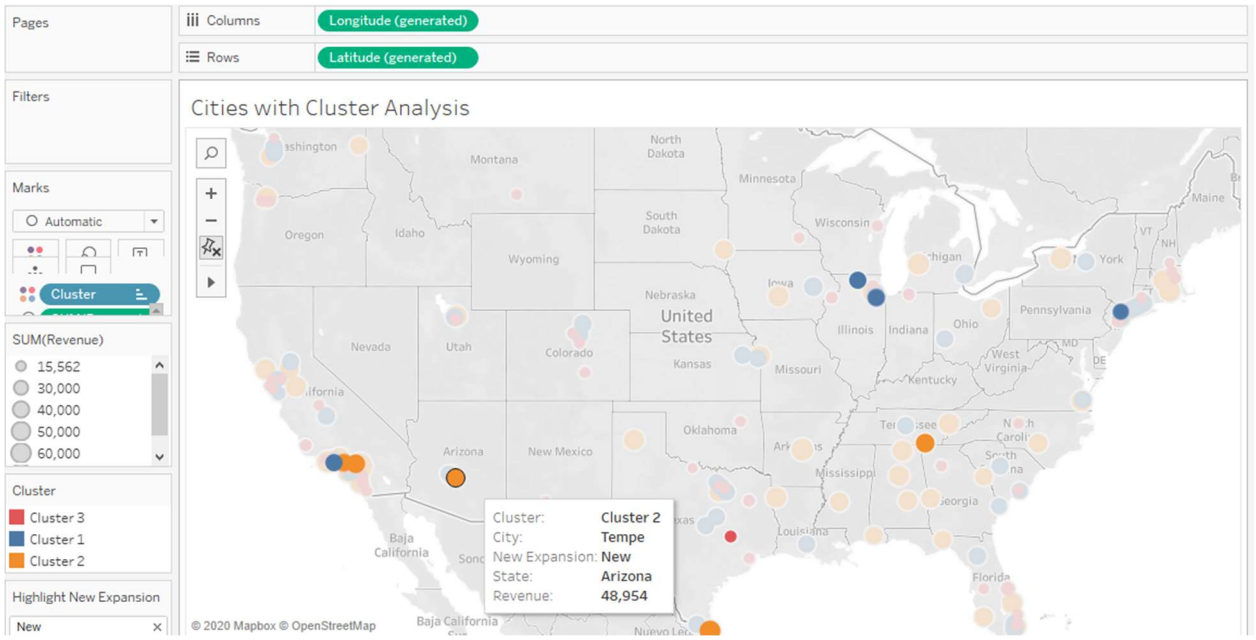
The clusters made earlier in the project are now used as a dimension so that we can visualize the average population of these clusters.

7. Cities with Cluster Analysis



Finally, we have used the cluster dimension and marked the cities which fall in the category of the clusters and have used the highlight function to highlight the cities accordingly.

8. New Cities with Cluster Analysis



So now we have the final new cities which contribute to the Returns on Investments. The cities in blue and Yellow have the highest returns on marketing investments.

The cities are as follows

1.Tempe, Arizona

2.Glendale, California

3. Rancho Cucamonga, California

4. Brownsville, Texas

5. Chattanooga, Tennessee

6.Joilet, Illinois

7.Rockford, Illinois

So out of 10 new cities, these cities will contribute to the return on marketing investment the most.

5. Deployment

After successful tableau representation, the report consisting of the findings would be handed over to the board of directors to help them make the decision regarding where to invest more funds for marketing.

Risk Factors

- Pandemic Outbreaks such as Covid-19
- Olde Census data of 2015
- Variance in State-wise laws and regulations

6. References

https://sds-platform-private.s3-us-east-2.amazonaws.com/uploads/P1-StartupExpansion.xlsx

https://sds-platform-private.s3-us-east-2.amazonaws.com/uploads/P1-US-Cities-Population.csv

7. Appendix

Shubham Srivastava

- Business understanding and the Data Preparation task for the project

Kavit Shah

- Evaluation and the tableau presentations for the project

Raj Mahajan

- Modeling and Deployment for the project

Thank You!