

VIDEO GAME POPULARITY DATA PROJECT

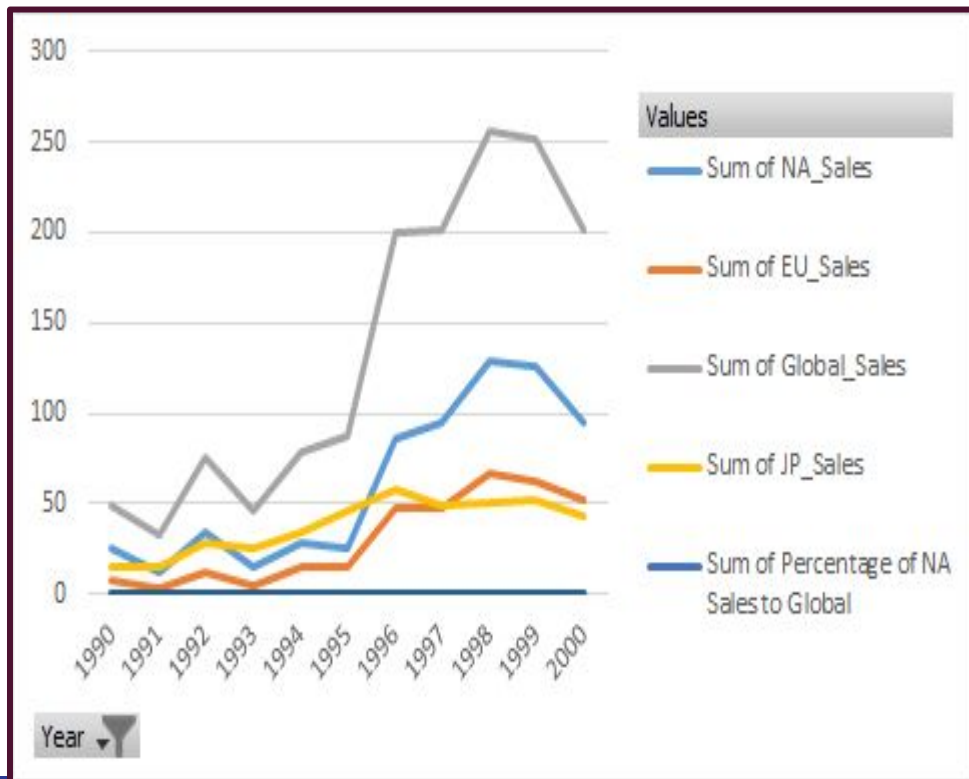
Descriptive analysis of video game data set for GameCo. a fictional game company, looking to improve marketing and sales of games for the next couple of years.

- **Are certain types of games more popular?**
- **Which publishers will be more competitive in certain markets?**
- **What games have increased in popularity over time?**
- **Sales figure variation between geographic regions.**

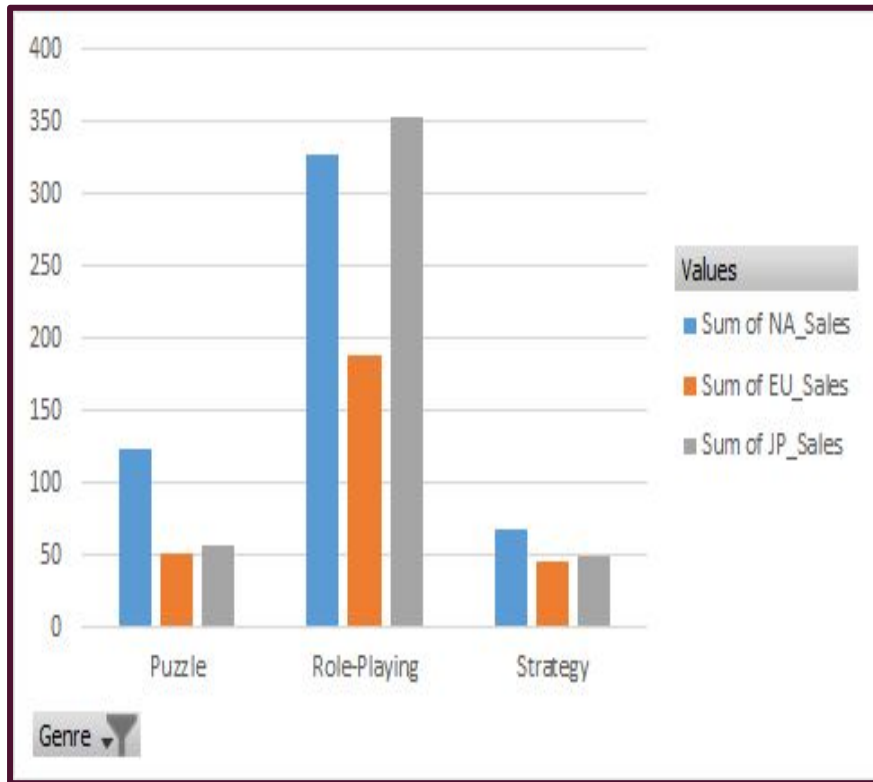
- **Grouping & summarizing data in Excel**
- **Descriptive analysis**
- **Visualizing results in excel**
- **Presentation for CFO, VP of marketing and senior VP of sales**



Sales in North America, Europe, Global & Japan for years 1990 to 2000



Sales on highest selling genres



Recommendations

- North American sales is overall higher and peaks around 2008. As for production NA should receive more shipment, followed by EU and then JP.
- During 1991,1993 and 1995 JP Sales exceeded NA Sales.
- Highest selling genres were puzzle, role playing and strategy among the countries with highest sales in US, Europe & Japan.
- Action, sports and shooter games are more popular in NA and EU while Role play , sports and action are popular in JP

CASE STUDY FOR A MEDICAL STAFFING AGENCY IN THE US

Determine when to send staff, and how many, to each state.

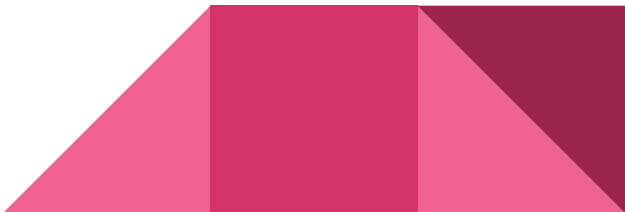
- Hospitals and clinics need additional staff to adequately treat extra patients. The medical staffing agency provides this temporary staff.
- The agency covers all hospitals in each of the 50 states of the United States, and the project planned for the upcoming influenza season on when and how many staff to each state optimally.

Data Sets:

- The following data sets covering influenza in the United States were used during the project: Influenza deaths by geography, time, age, and gender from CDC, Population data by geography Source from US Census Bureau, Counts of influenza laboratory test results by state (survey) from CDC, Influenza Visits Data Set and Lab Tests & Survey of flu shot rates in children from CDC.

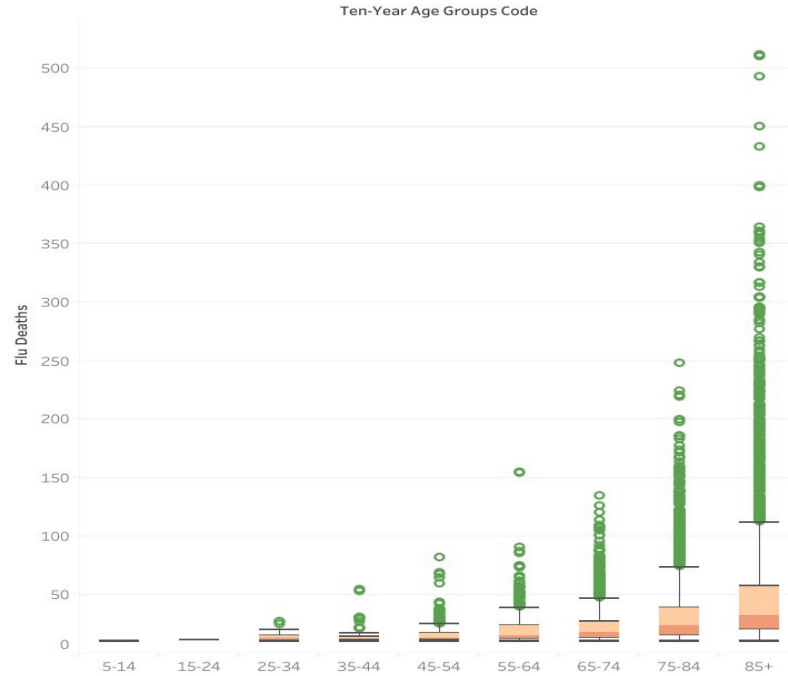


Analysis conducted:

- **Designed & formulated research project**
 - **Data cleaning & summary statistics**
 - **Selection & Integration of data sets**
 - **Identified and create hypothesis, then tested with data**
 - **Relationship between variables & correlation**
 - **Spatial and temporal visualization**
 - **Communicate recommendations & next steps based on research findings**
 - **Published on tableau**
- 

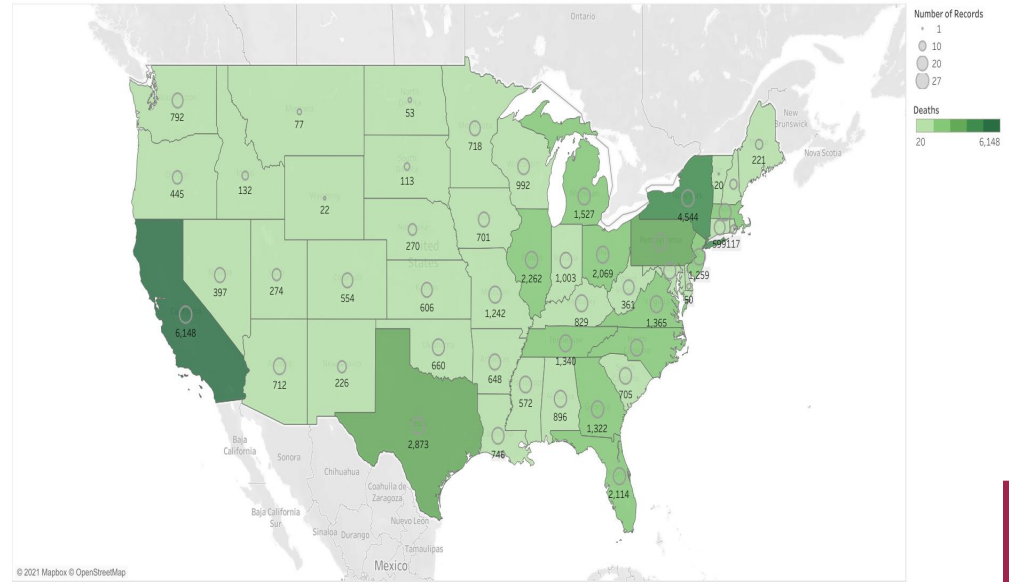
Flu death by age-group

Box and Whisker Plot Flu Deaths



States with highest flu deaths.

FludeathMonthTop5



Recommendations

- California, New York, Texas, Florida and Illinois have the most flu related hospitalizations and death during the years of 2011 through 2017. Staffing should be high priority for these states.
- Flu season is highest during the months of January, March, February and December. Staffing needs should be met for the highest need during the period.
- Age groups 65 and over should be encouraged to vaccinate to prevent illness in all states
- The drop in flu deaths from 2015 to 2016 could be further analyzed and applied to the upcoming years.

Link to the tableau storyboard:

<https://public.tableau.com/app/profile/uma.chidambaram/viz/StoryPlanForAgency/StoryPlanForAgency>



MOVIE RENTAL COMPANY ROCKBUSTER: LAUNCH OF AN ONLINE VIDEO RENTAL SERVICE

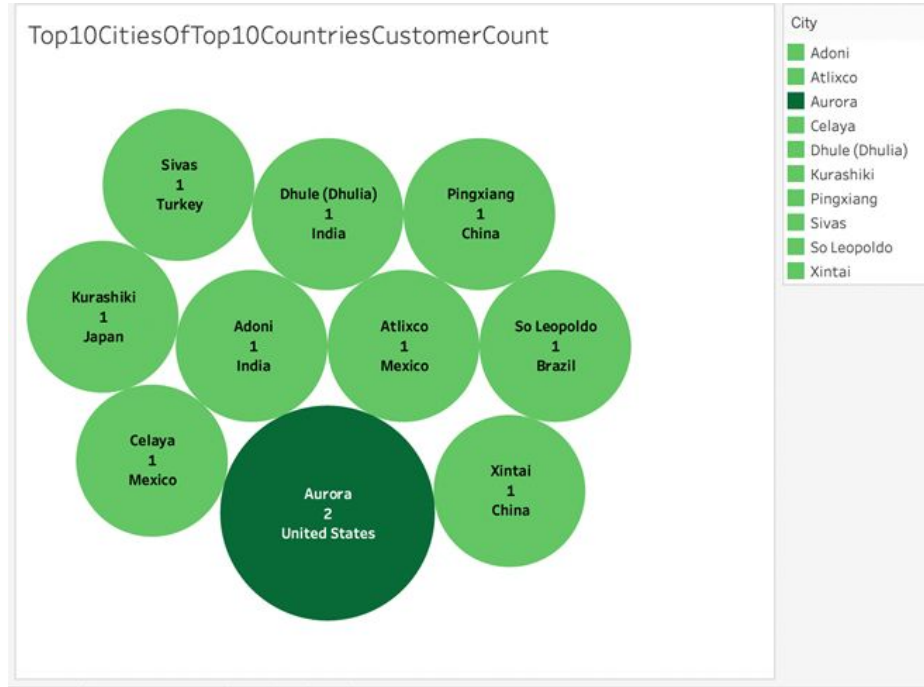
As a business analyst, develop customer insights and launch strategy.

- Loaded Rockbuster's data into a relational database management system (RDBMS).
- Used SQL to analyze the data and answer any ad-hoc business questions that other departments may have.
- Filter, clean, summarize, join tables and common table expressions for desired results.

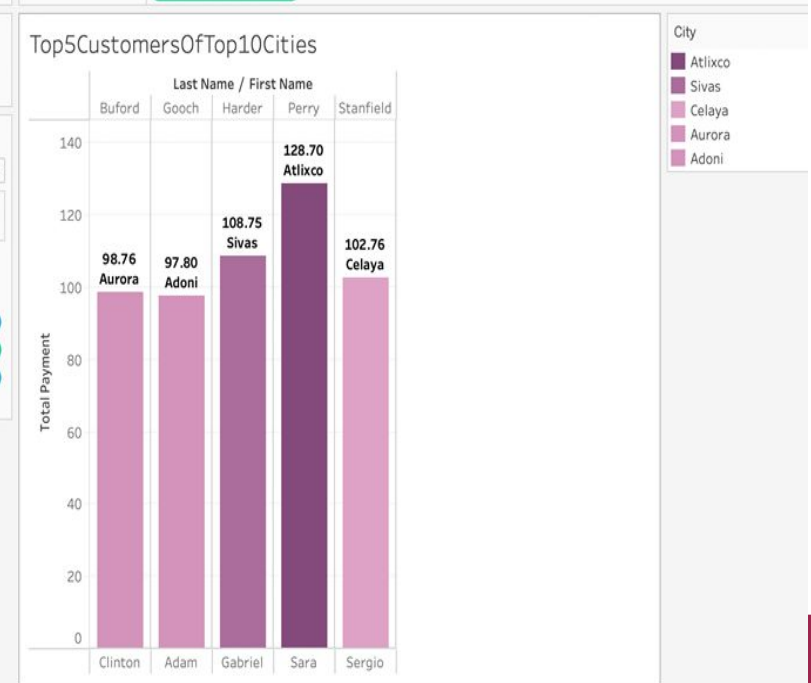
- Movies that generated most/least revenue.
- Average rental duration & countries customers are based in.
- Sales figures from different geographic regions & customers with a high lifetime value based and their location.



Cities with highest rentals among the countries with highest rentals.



Loyal Customers with highest rental.



Recommendations:

- Analysis on top 10 countries with highest customers are India, China, United States, Japan, Mexico, Brazil, Russian Federation, Philippines, Turkey and Indonesia.
- Top ten cities of the top ten countries Aurora(US), PingXiang(China), So Leopoldo(Brazil), Celaya(Mexico), Sivas(Turkey), Dhulia(India), Kurashiki(Japan), Atlixco(Mexico), Xintai(China) and Adoni(India).
- Top ten customers that were to be rewarded for the most loyal customers are Buford Clinton, Gooch Adam, Harder Gabriel, Perry Sara and Stanfield Sergio.

<https://github.com/UmaChidu>

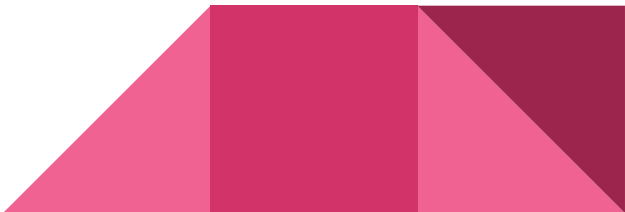


INSTACART, ONLINE GROCERY STORE:

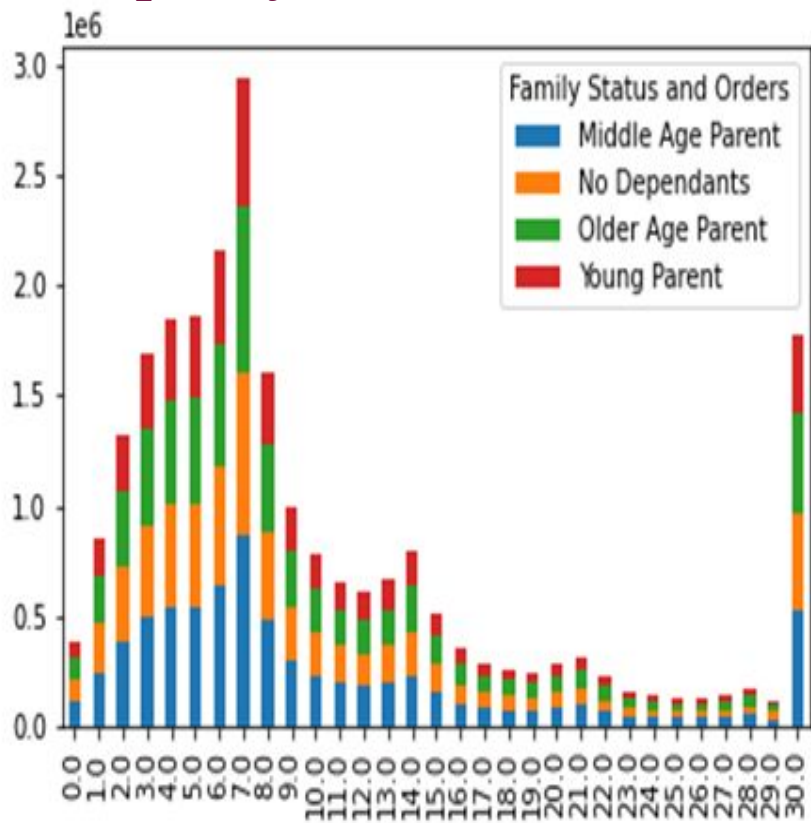
Uncover sales patterns for targeted marketing

- Sales team needs to know busiest days of week and hours of day and times of day when customers spend most money.
- Marketing and sales want to use simpler price range and popular products and frequency of orders and groupings to help direct their efforts.
- Creating customer profile by brand loyalty, customer region, loyalty, ordering habits, demographic, family, age, ordering habits and regional status.

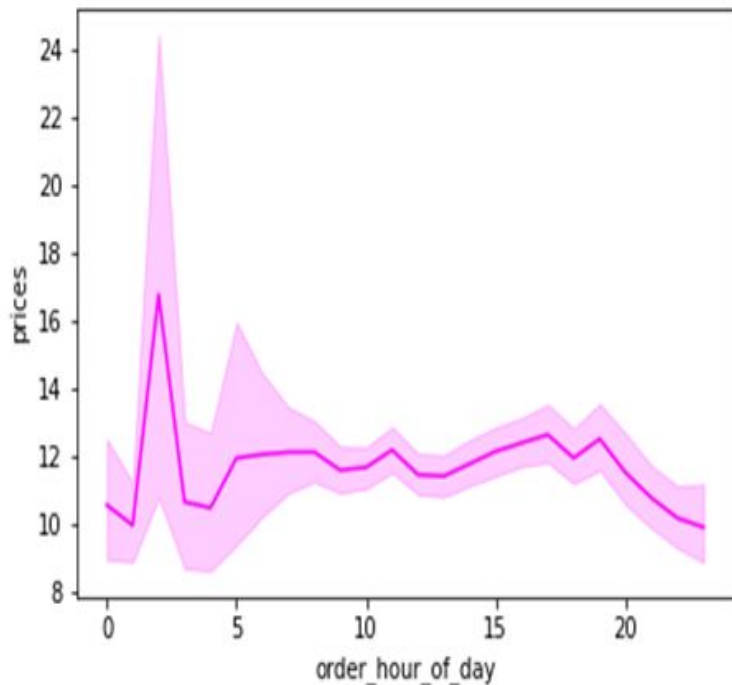


- **Analysis conducted on python and relevant libraries (pandas, NumPy, os, matplotlib, scipy, and seaborn).**
 - **All required data sets have been successfully installed and imported into each script.**
 - **Data has been cleaned. Duplicate data, missing data, and mixed-type columns have been checked and addressed.**
 - **Data sets are merged into single data set. Merged data contains variables needed for analysis.**
 - **Required columns derived and relevant visualizations created.**
- 

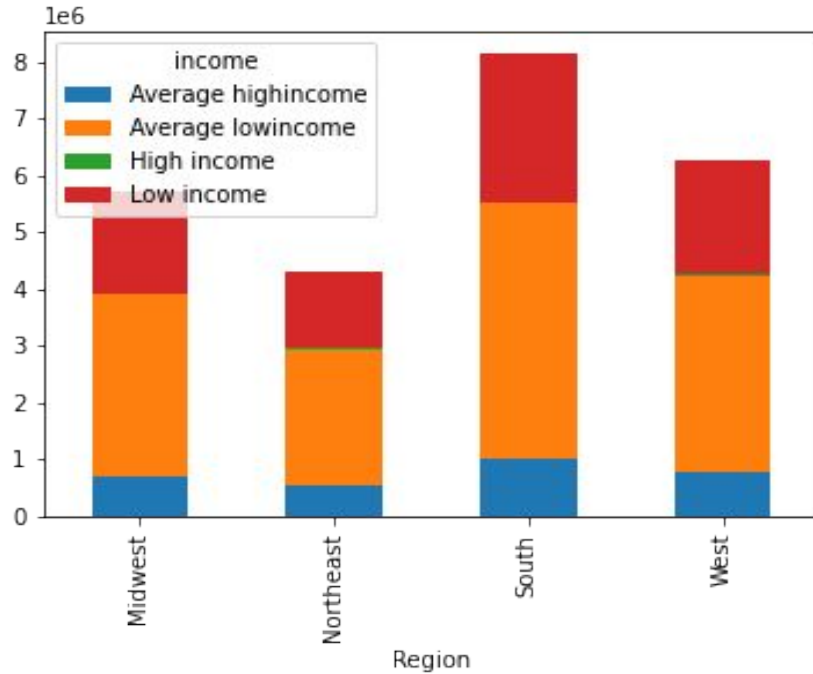
Family status and daily frequency of orders.



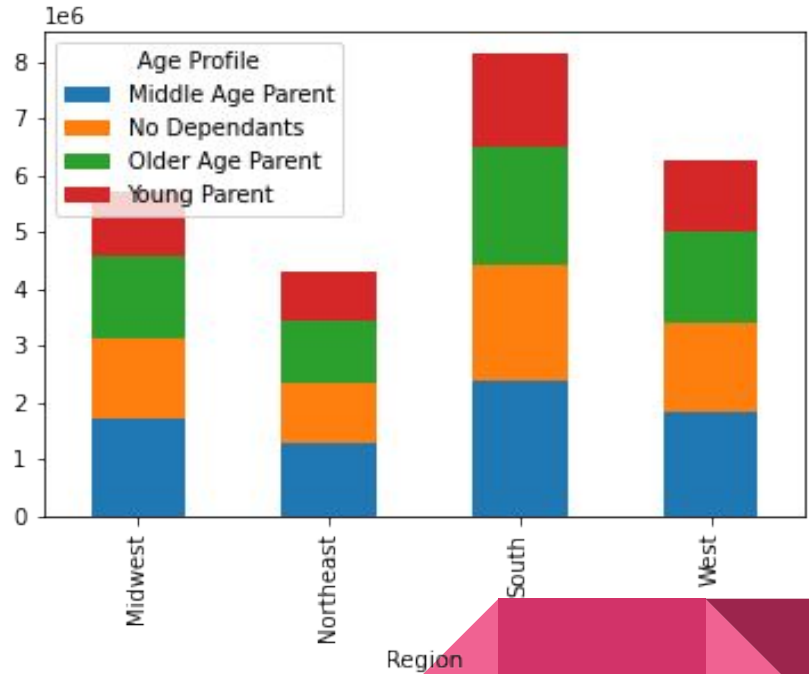
Orders time of day



Income category in Regions



Family Status and Region



Recommendations

- Highest orders are placed within 7 days of prior order. Users stay loyal to instacart from this for grocers and perishables. There is also a big increase in 30 days which could be household items purchases.
- 33% of customers from South and 26% are from the West are middle aged parents. 33.4% of customers from South and 26% from the West care for older aged parents .
- 29% of loyal customers are middle aged parent. 25% of them care for older aged parent and another 15% have no dependents.
- Middle aged parents (child as dependent), older age parent (as dependants) and customers with no dependents make more orders as opposed to young parents.

DATA ANALYSIS FOR A FICTIONAL GLOBAL BANK

Anti-money laundering project at the bank.

- **Big data and extracting knowledge from big data**
- **Data ethics**
- **Data mining and decision making**
- **Predictive analytics and models using linear regression**
- **Time-series analysis & Forecasting**
- **Github to collaborate with colleagues**



Not Exited

Country	Count of Gender
France	403
Germany	182
Spain	202
Total	787

Exited

Country	Count of Gender	Percent Excited
France	77	16
Germany	75	29
Spain	52	20
Total	204	21

Age	Balance	Salary	Credit Score
37	74831	98943	652

Average of Not
Exited

Age	Balance	Salary	Credit Score
45	90239	97155	637

Average of
Exited

Recommendations

- Main reasons customer exit bank are age, salary, gender, country of residence and credit score.



STOCK EXCHANGE ANALYSIS

Data Immersion: Exploratory Analysis from sourcing data.

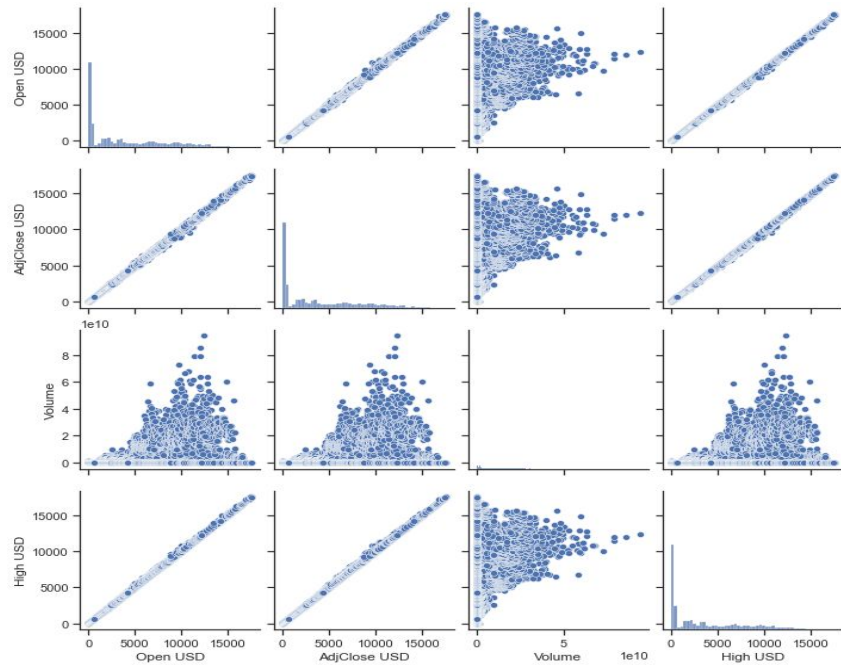
Stock exchange data from kaggle, yahoo finance <https://www.kaggle.com/mattiuzc/stock-exchange-data> is used for analysis.

Daily price data for indexes tracking stock exchanges from all over the world (United States, China, Canada, Germany, Japan, and more).

Prices in terms of the national currency of where each exchange is located.

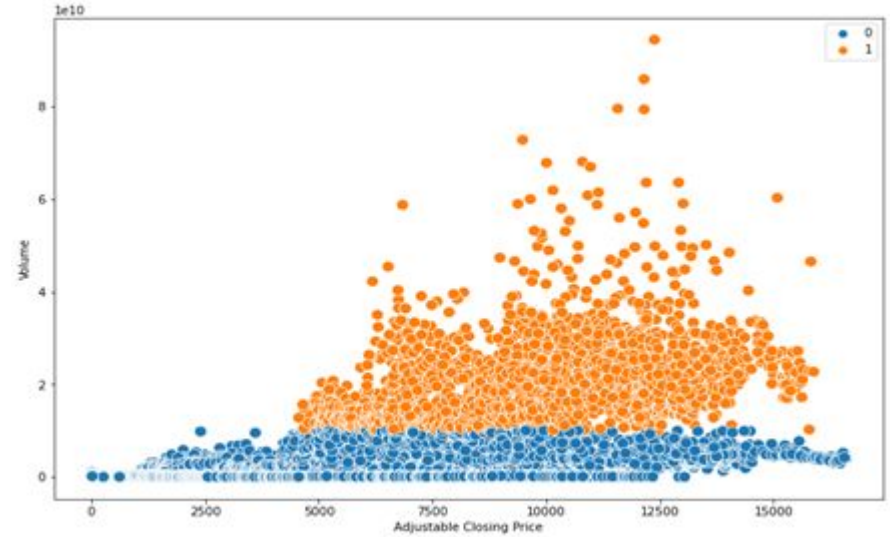
To conduct initial visual analysis, I used python and its libraries.





After cleaning the data, created paired plots to analyze the relationship between the columns.

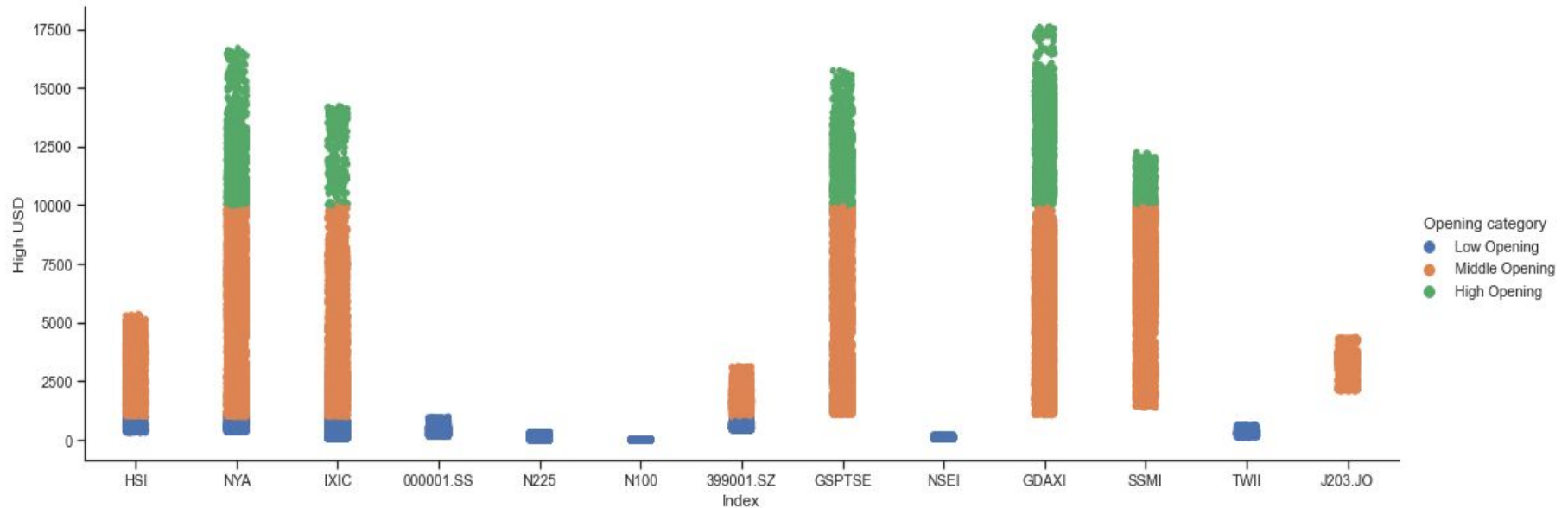
Volume is more around the middle or when the Opening is between 5000 and 15000 USD.



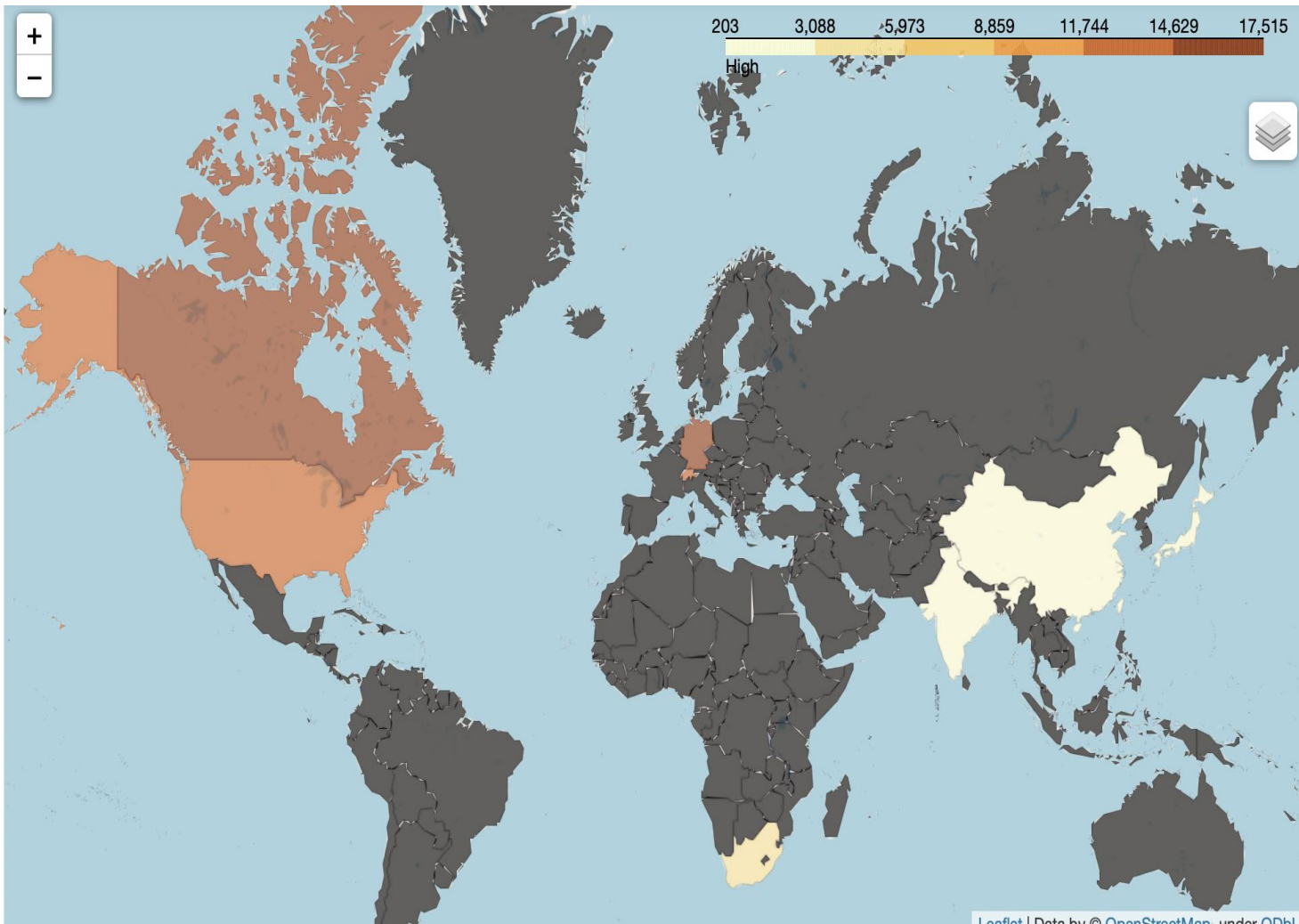
After using machine learning and K-means algorithm to determine optimum number of clusters for analysis created the visual comparing Open and adjustable close with Volume.

Adjustable Closing Price of 10000 to 12500 USD tend to have more Volume with a peak at 12500. Volume is higher for the orange category.

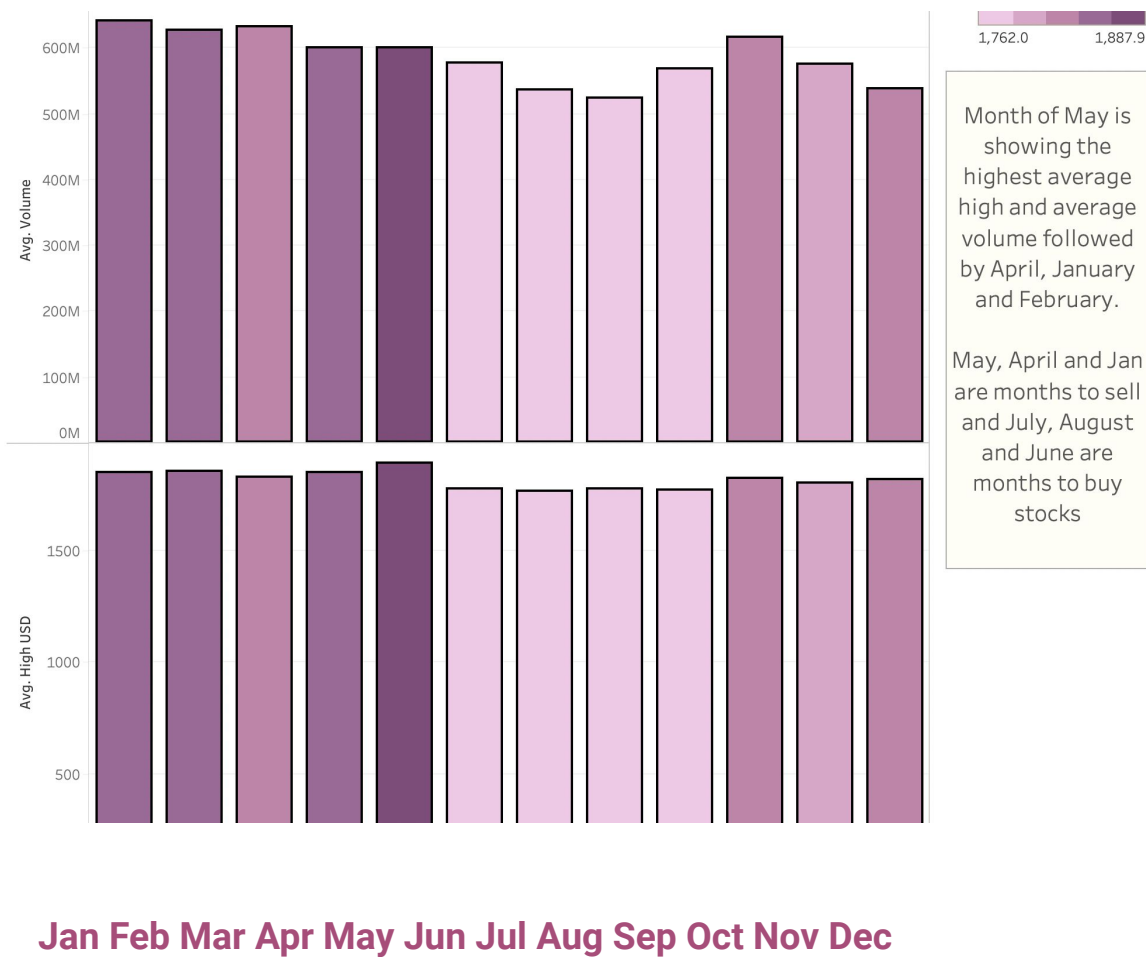
Categorical plot of each stock in the data yield the following:



US, Switzerland, and Canada are the countries with highest selling stocks. With global trading and picking out the right index we can be on track for wise investing.



Conducted geospatial analysis using geojson using location data by creating a choropleth map. USA, Canada, Germany, China, India, Johannesburg and Taiwan have the Highs. Stocks from countries of dark brown shade to lighter are safer bets.



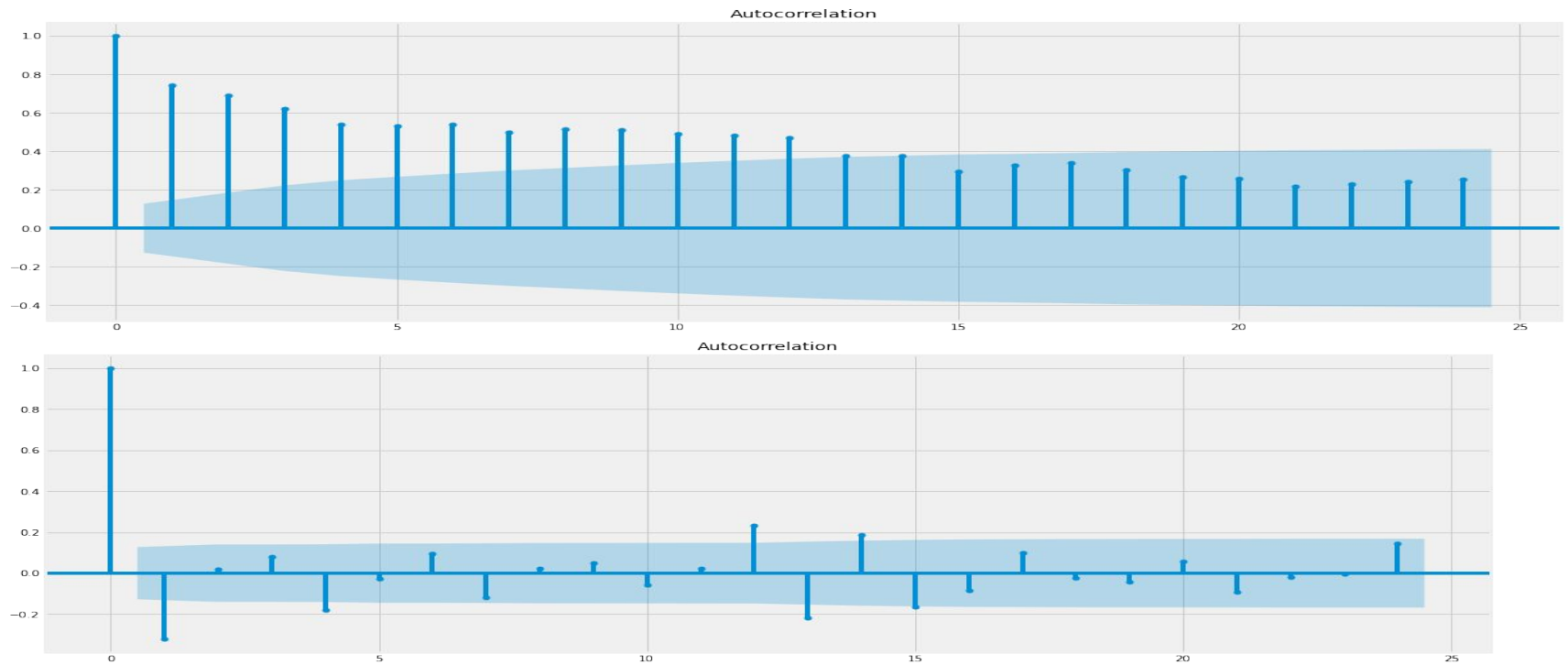
- **After grouping by month on Volume and average high got the visualization on tableau.**
- **Monthly average volume and average high show highest for April, January and February.**
- **May, April and January are months to sell**
- **July, August and June are months to buy stocks to be profitable.**



To plot the time series analysis of the data, I grouped Volumes of each month and build a dataframe with data for each month and only year after 2000. There is a peak during the year 2009 and 2021. This can be explained by the subprime crisis that led to higher openings in 2007 and the intervention of federal reserve in 2008 to rescue bear market.

The 2001 recession was relatively mild after the tech bubble burst and was worsened by the 911 tragedy.

We can see the peak in 2009 and 2021. The dip in 2013 was due to government shutdown and the battle between house and senate on a healthcare bill.



The Dickey-Fuller test was performed to test the stationarity of data to check whether they are ready for predictions. Differencing operation was performed to stationarize the data. The top shows before and bottom shows after differentiation and stabilized data.

Congress continues to attempt to make the market a safer place for individual investors by empowering the SEC.

One example of this is the Sarbanes-Oxley Act (2002). After Enron, WorldCom, and Tyco International used slippery accounting that resulted in widespread damage to investor portfolios, the SEC was given the responsibility to prevent a repeat in the future.

In addition to short term checks on stocks it is evident that world events relating to recession and its causes should be kept in check for gaining from investment in stocks. For long term investments, in addition to watching for world events, it's safer to invest in stocks that consistently perform well over the course of years, like NYA and HSI.

The Jupyter scripts are available [here](#) in my GitHub repository.

An interactive dashboard can be accessed in [Tableau](#).

by

Uma Chidambaram.

Contact me at uma.chidu2001@gmail.com.

