Katie Kulp

Data Analyst Portfolio

2023

Projects and Relevant Skills

➢ *1. Preparation for Influenza Season*	- *Excel* - *Tableau* - *Effective Communication of Results*
➢ *2. Transition of RockBuster to Digital Platform*	- *SQL* - *Database Dictionary* - *Powerpoint Presentation*
➢ *3. Optimizing Customer Segmentation for Instacart*	- *Python* - *Jupyter Notebooks* - *Python Visualizations*

Project 1: Preparation for Influenza Season

Project Overview :

Working with a medical staffing agency that provides temporary workers to clinics and hospitals.

The analysis involved the examination of historic influenza trends to anticipate where, when, and how many staff will be needed to minimize deaths attributable to the flu.

Data:

The data came from US Census Bureau and the CDC.

Two data sets were cleaned and combined.

Project 1: Preparation for Influenza Season

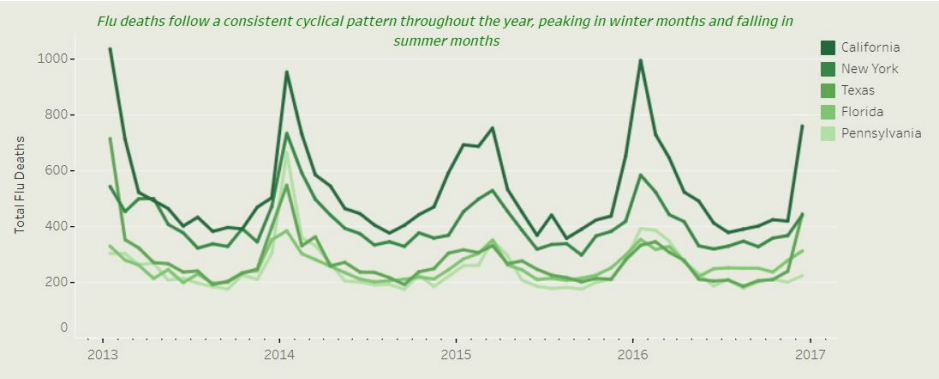
The data was cleaned and missing, duplicate, or erroneous values were managed. The two data sets were combined on a common key.

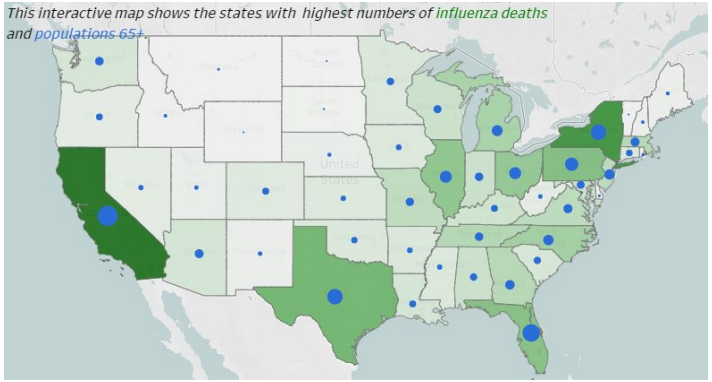
Statistical analysis was performed in excel with basic descriptive analysis and a one tailed t-test.

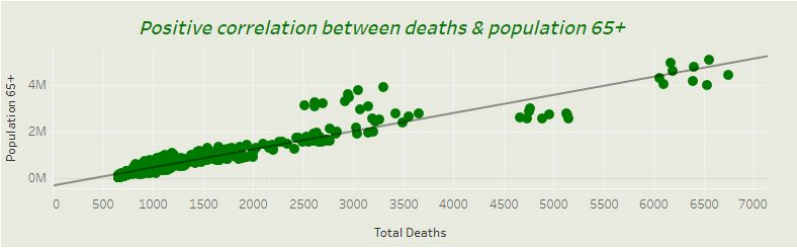
Datasets were imported in to tableau where visualizations were made to reflect the relationships between variables.

The final results were polished and combined into a cohesive storyboard presentation.

Summary: With the intent of determining when and where additional assistance for the upcoming flu season, it was found the most at risk are elderly populations (ages 65+) and those in populous states, especially during the winter months







Link to tableau presentation: https://public.tableau.com/app/profile/katie.kulp/viz/A2_storyboard/Story1

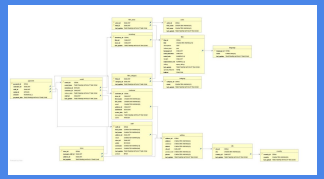
Project 1: Preparation for Influenza Season

Project 2: Transition of RockBuster to Digital Platform

Project Overview :

The project was working with a movie rental store that was looking to transition to streaming platform.

They wanted to target marketing by understanding which movies, customers groups, and locations contributed most to revenue.



Data:

The data from Rockbuster consisted of several folders with information on inventory, customers, payments and was loaded into PostgreSQL database.

Project 2: Transition of RockBuster to Digital Platform

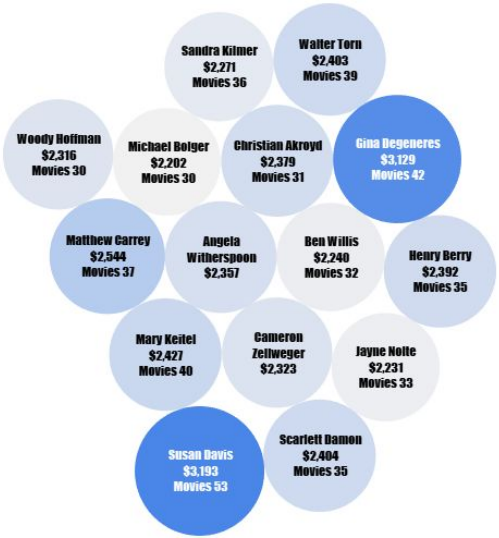
The data was downloaded and the database environment set up using PostgreSQL.

Once in the database, the data was queried using SQL in order to find results for the companies main questions.

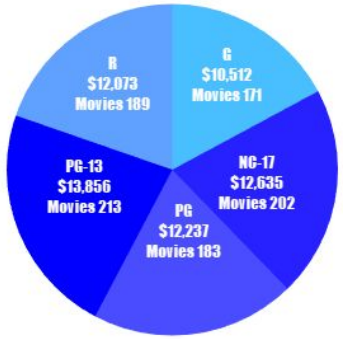
The results of the query and the relationships between the variables were visualized using Tableau.

The final results and recommendations were collected and prepared in a powerpoint presentation.

Summary: With the goal of determining how best to set up their digital movie platform, it was found that Rockbuster would benefit most from a diverse selection of movies, loyalty incentive programs, and a focus on several high revenue countries.

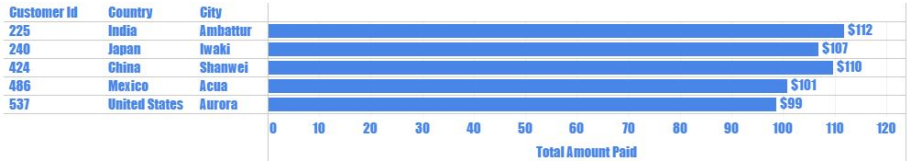


A diagram of the highest grossing actors for Rockbuster



The movie ratings that generated the most revenue and the count of movies for each rating

The highest grossing movie genres for Rockbuster and count of inventory for each



The locations of customers that generated the most revenue for Rockbuster

Link to GitHub Repository : https://github.com/kkulp-lab/SQL_Rockbuster_Analysis/tree/main

Project 3. Optimizing Customer Segmentation for Instacart

Project Overview :

This project was working with Instacart with the intention of analyzing sales trend to better target marketing.

They were focused on customer demographics, segmentation, and time and day of highest usage.

Data:

The data consisted of several data sets on customers, orders, and products. They were combined using python into one data set (32404859 rows).

Project 3. Optimizing Customer Segmentation for Instacart

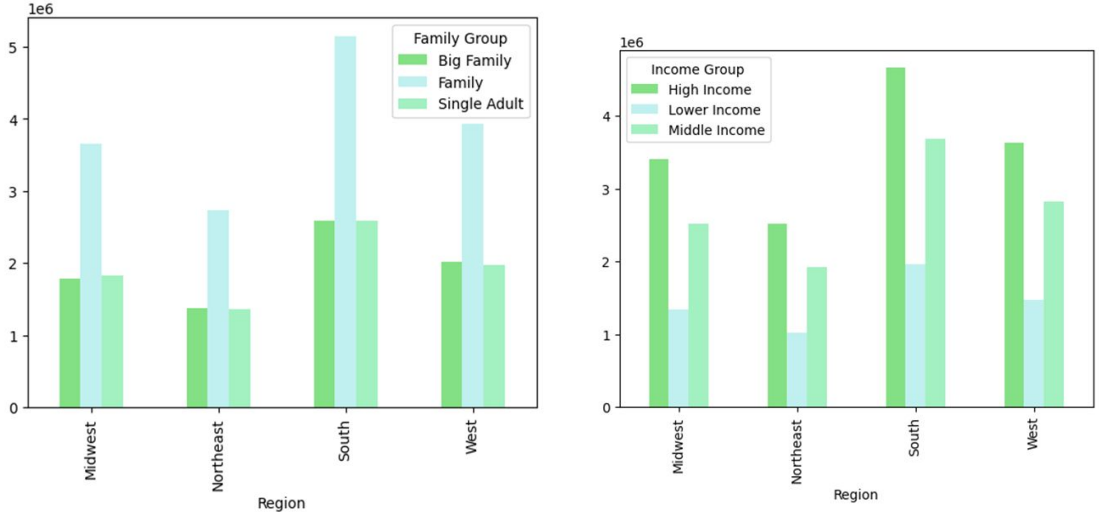
Data was cleaned using python. Changes included deleting, renaming columns, finding duplicate and missing data.

Statistics were performed using grouping and aggregating, and new variables and columns were derived to analyze customers in different groups.

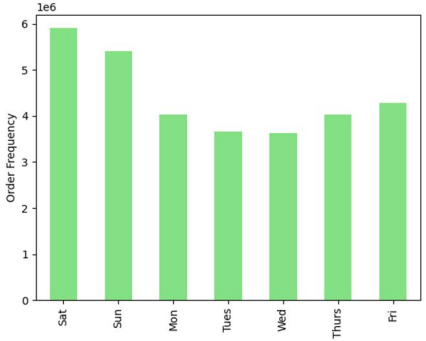
The results were visualized using python on jupyter notebook and exported into the final excel report.

A final excel report was made with documentation on the data cleaning process as well as final results and recommendations.

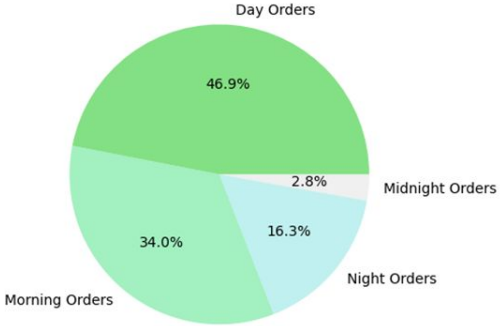
Summary: It was found that across all regions most orders came from higher income, two parents households with children during the weekend or on Friday and most commonly in the morning or daytime.



Bar charts showing that the most customer orders came from high income families, regions and customer groups are derived variables



A histogram displaying DOW with most orders



A pie chart showing the time of day with most orders, groups are derived variables

Link to GitHub Repository : https://github.com/kkulp-lab/Python_Instacart_Analysis

Project 4. California Communities Analysis

Project Overview :

This project was an in depth analysis of the relationships between various characteristics of CA counties.

The goal was to find any relationships between demographic and health markers that could predict which communities may suffer from poor health and need additional assistance.

Data:

The data comes from the California Environmental Protection Agency. It consists of various environmental, demographic, and health markers by CA census tracts.

Project 4. California Communities Analysis

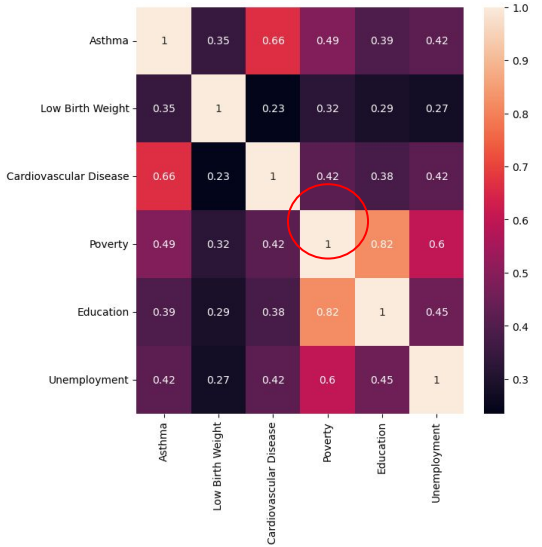
Data was cleaned using excel and python. Changes included removing spaces, renaming and dropping columns for use.

Several methods of relationship analysis were performed including time series, linear regression, and cluster analysis.

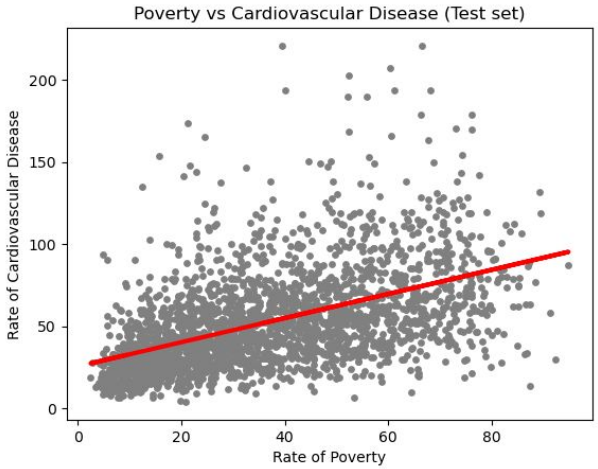
The resulting visualizations from the analysis were recreated on Tableau.

An interactive Tableau storyboard was the final report showing the main points of the analysis and final results and recommendations.

Summary: It was found that there is a weak relationship between poverty and cardiovascular disease. To mediate, more resources, education, and healthier food and environments should be available to those who are living with lower income.

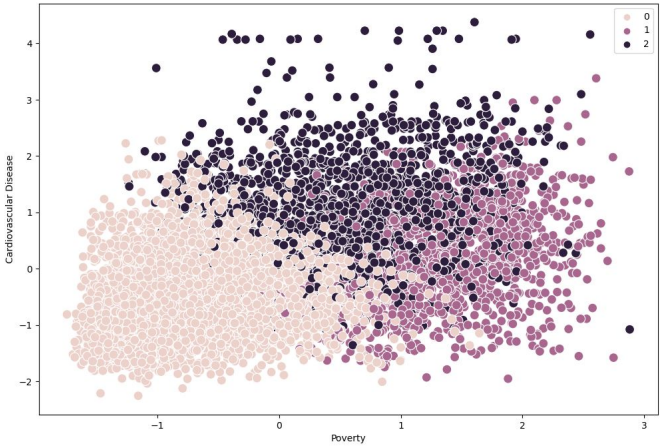


Exploratory analysis indicated there may be a weak relationship between cardiovascular disease and poverty



Supervised machine learning linear regression was performed on the data set. The results indicated that about 24% of cardiovascular disease instances could be predicted by poverty markers.

A cluster analysis was then performed on the data in attempt to further identify relationships between variables. In general, lower poverty rates were clustered with lower rates of cardiovascular disease.



Link to Tableau Storyboard : https://public.tableau.com/app/profile/katie.kulp/viz/Exercise_6_7/Story1?publish=yes

Link to GitHub Repository : https://github.com/kkulp-lab/Python_California_Analysis

Please reach out for any questions...
katie_kulp@yahoo.com

https://github.com/kkulp-lab