

Christopher Evangelista

Data Analytics Project Portfolio



cnevangelista13@gmail.com



github.com/ChrisEvangelista1



public.tableau.com/app/profile/christopher.evangelista/vizzes



Projects



NBA Shot Analysis

Analyzing NBA shot efficiency by area of the court

Citi Bike Demand Analysis

Using Python / Streamlit to create a dashboard to analyze Citibike demand



FoodMart Data Reporting

Overhauling reporting for sales and finance teams through Tableau



Instacart Market Strategy

Generating insights through market segmentation analysis



Rockbuster Stealth, Inc

Creating a business plan for launching a streaming service



Pig E. Bank

Forecasting and analyzing retention of banking customers



Preparing for Influenza Season

Planning staff deployment through analyzing historical data



NBA Shot Analysis



Background

Shot analysis is essential to maximizing a team's offensive strategy and overall success. It helps players and teams identify areas of strength, uncover weaknesses, and optimize strategy.



Objectives

This project analyzes the 2023-2024 NBA season to understand the various variables such as shot selection, shot efficiency, and shot quality that makes a good NBA shooter.



Key Skills

- Tableau – data visualization
- Python – data wrangling and analysis



Key Questions

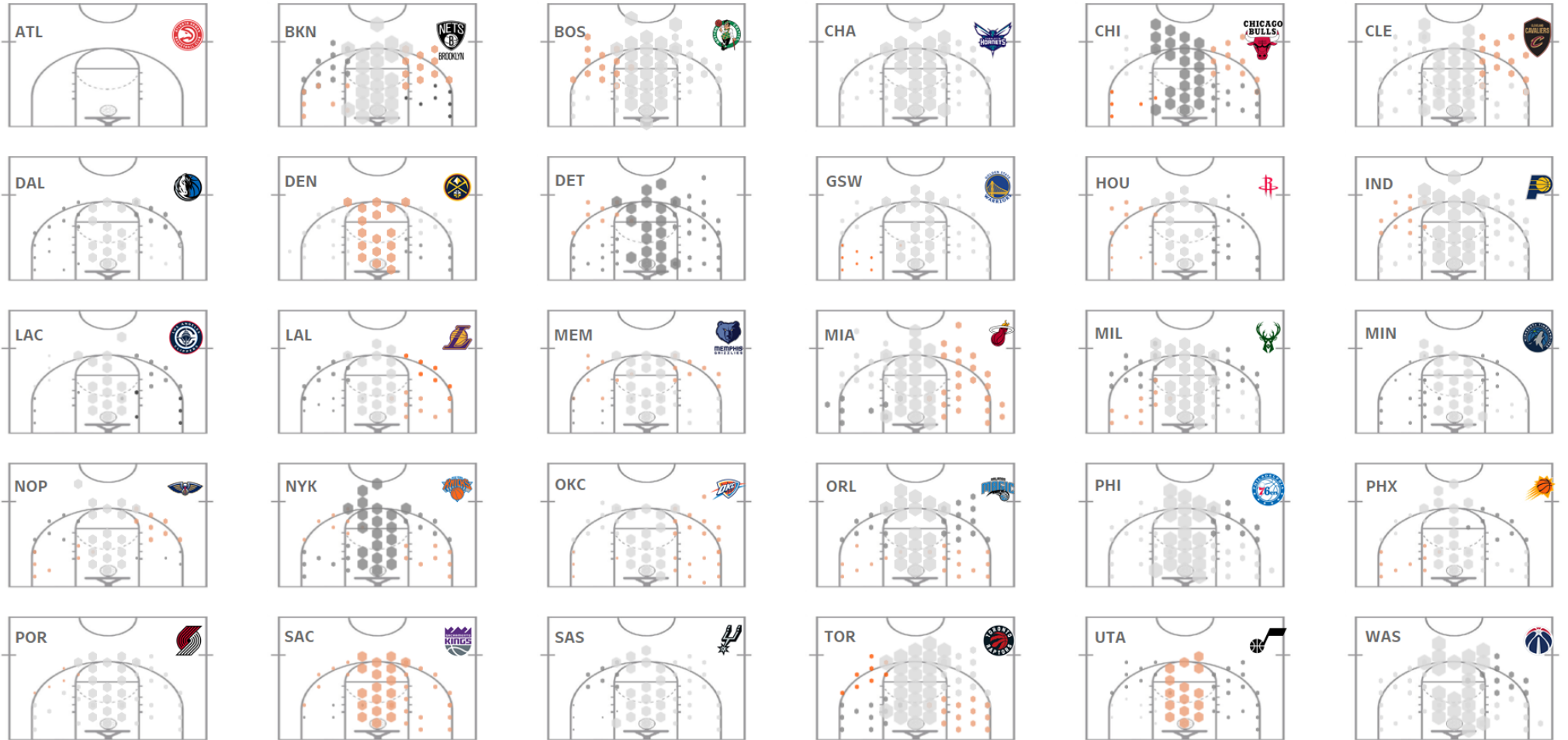
- Is there a positive correlation between the number of three-point attempts and overall field goal percentage?
- Is the expected value of a 3-pointer less than the expected value of a 2-pointer?

2023 - 2024 NBA Matchup Shot Analysis

Selected Team:

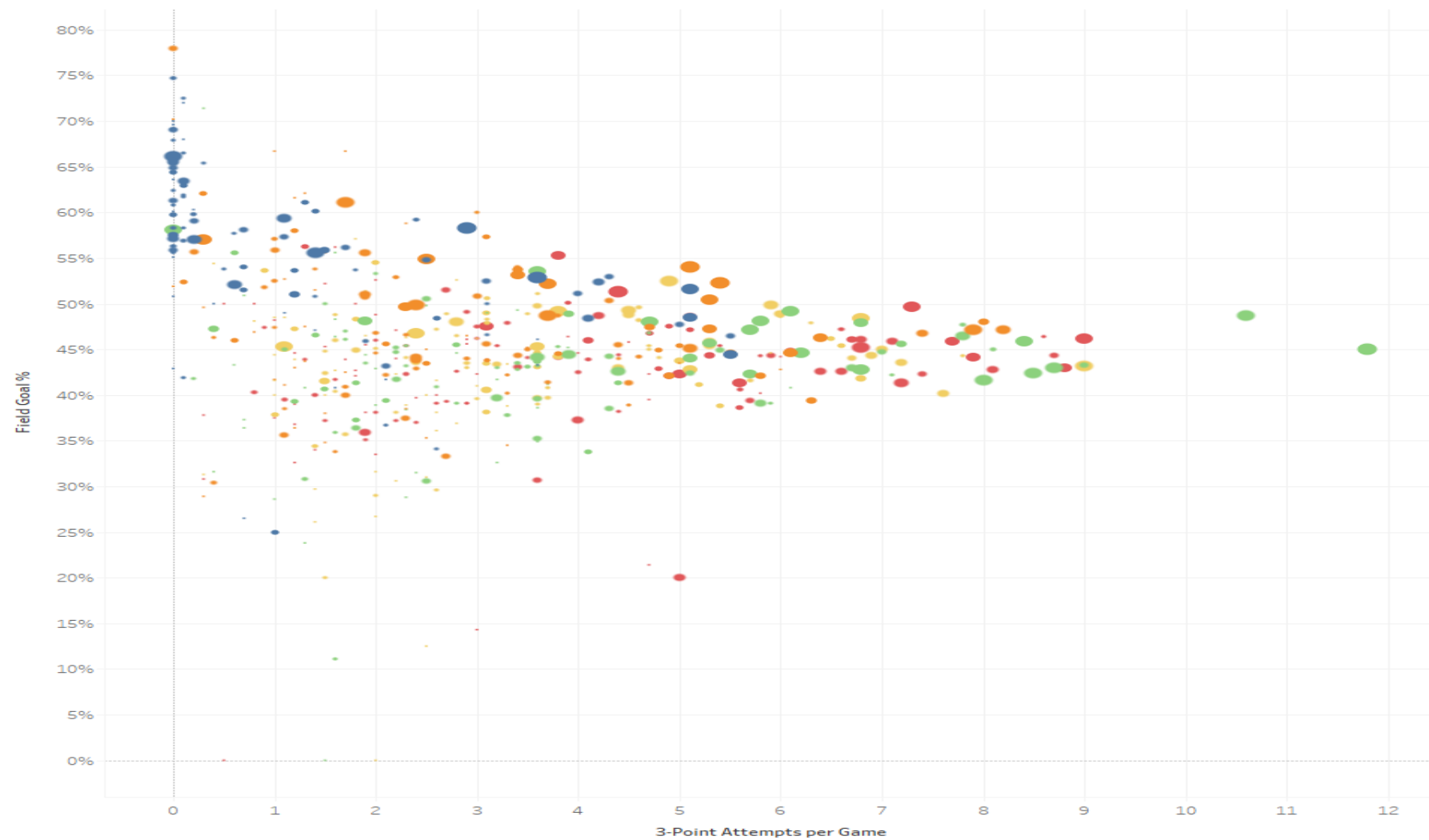
ATL

Below Average ●●●●● Above Average | Fewer Shots ●●● More Shots



Explore on Tableau Public

Three-Point Attempts vs Field Goal %





Citi Bike Demand Analysis



Background

To diagnose where distribution issues stem from and advise higher management on a solution based on your diagnosis of the root of the problem—whether it's sheer numbers, seasonal demand, or something else.



Objectives

The project's objective is to conduct a descriptive analysis of existing data and discover actionable insights for the business strategy team to help make informed decisions that will circumvent availability issues and ensure the company's position as a leader in eco-friendly transportation solutions in the city.



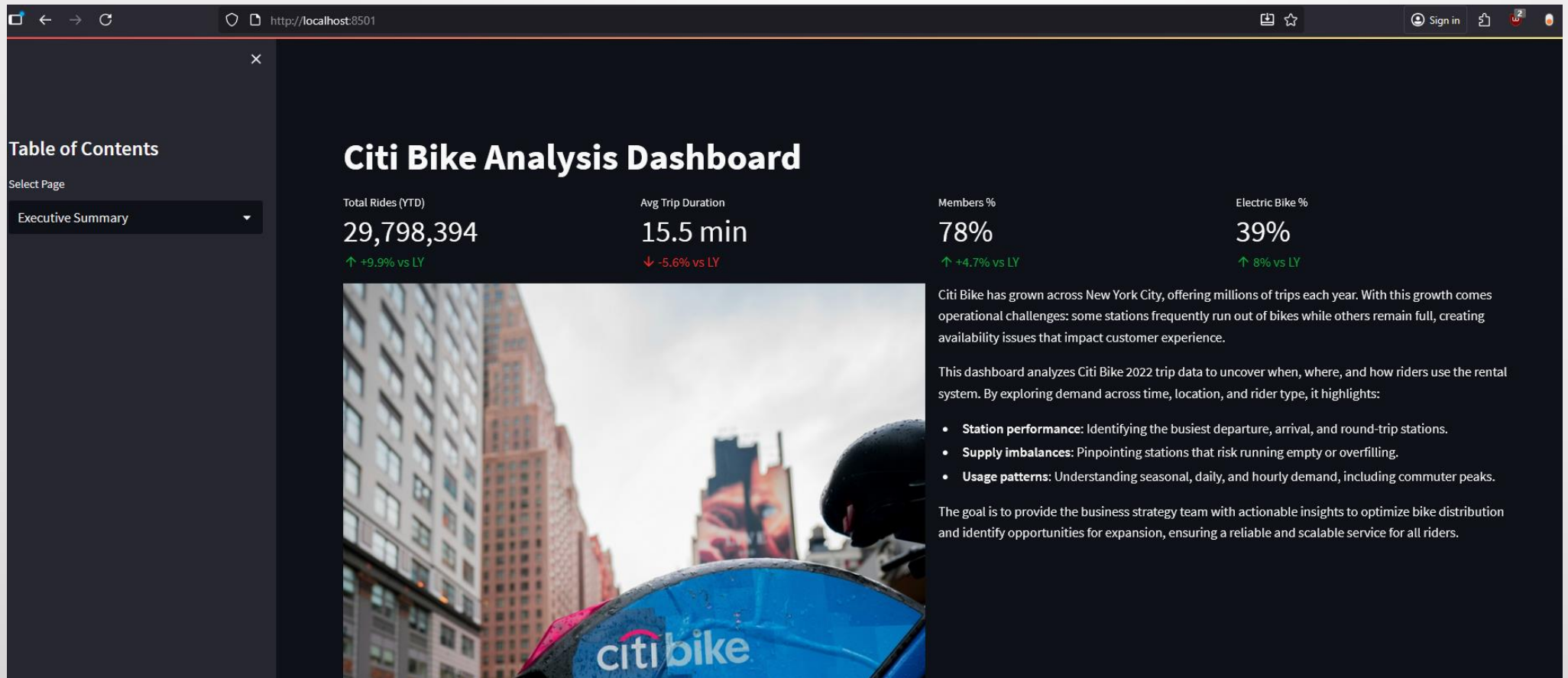
Key Skills

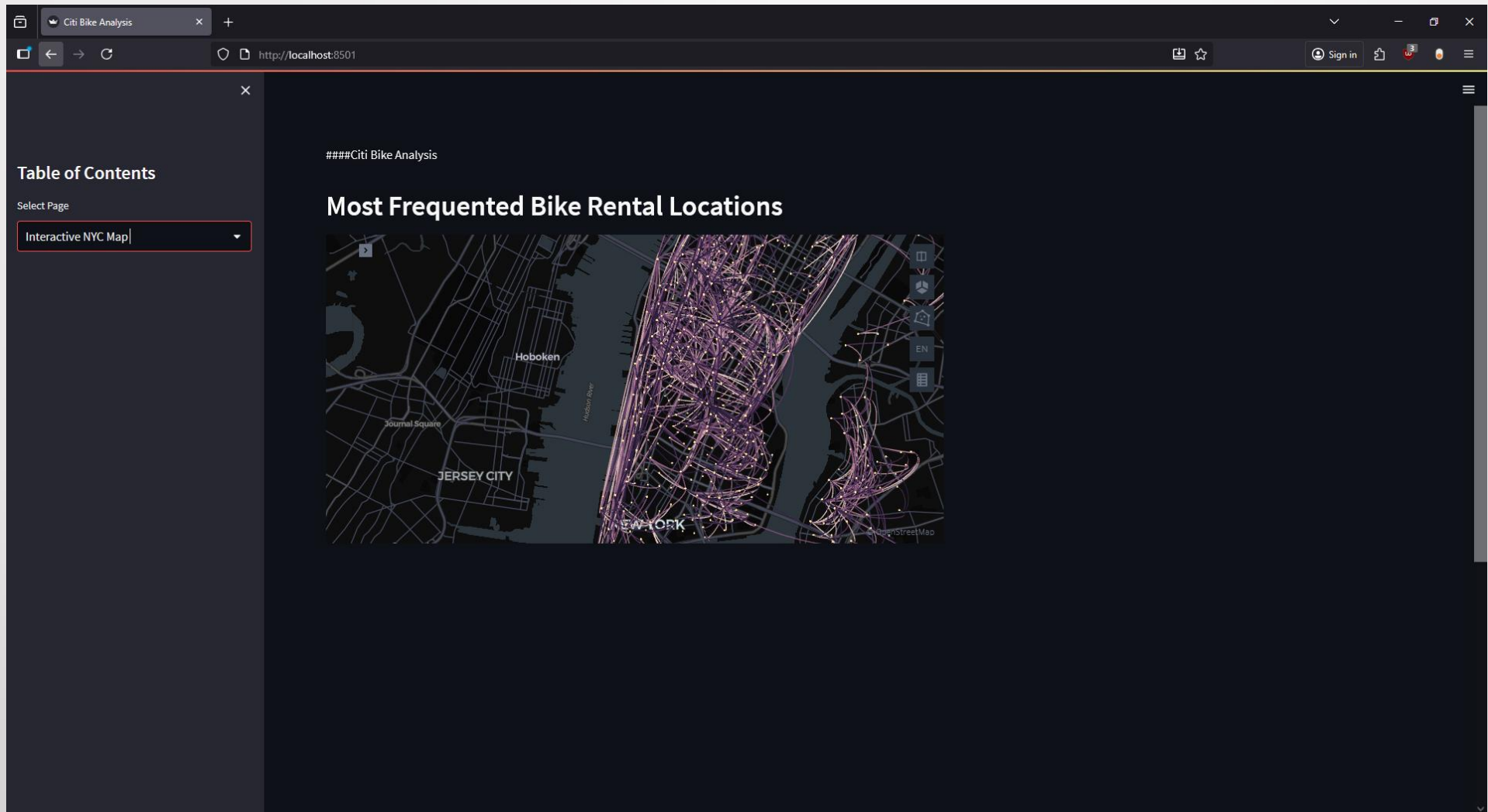
- Python
- Streamlit - dashboard
- Pandas - data

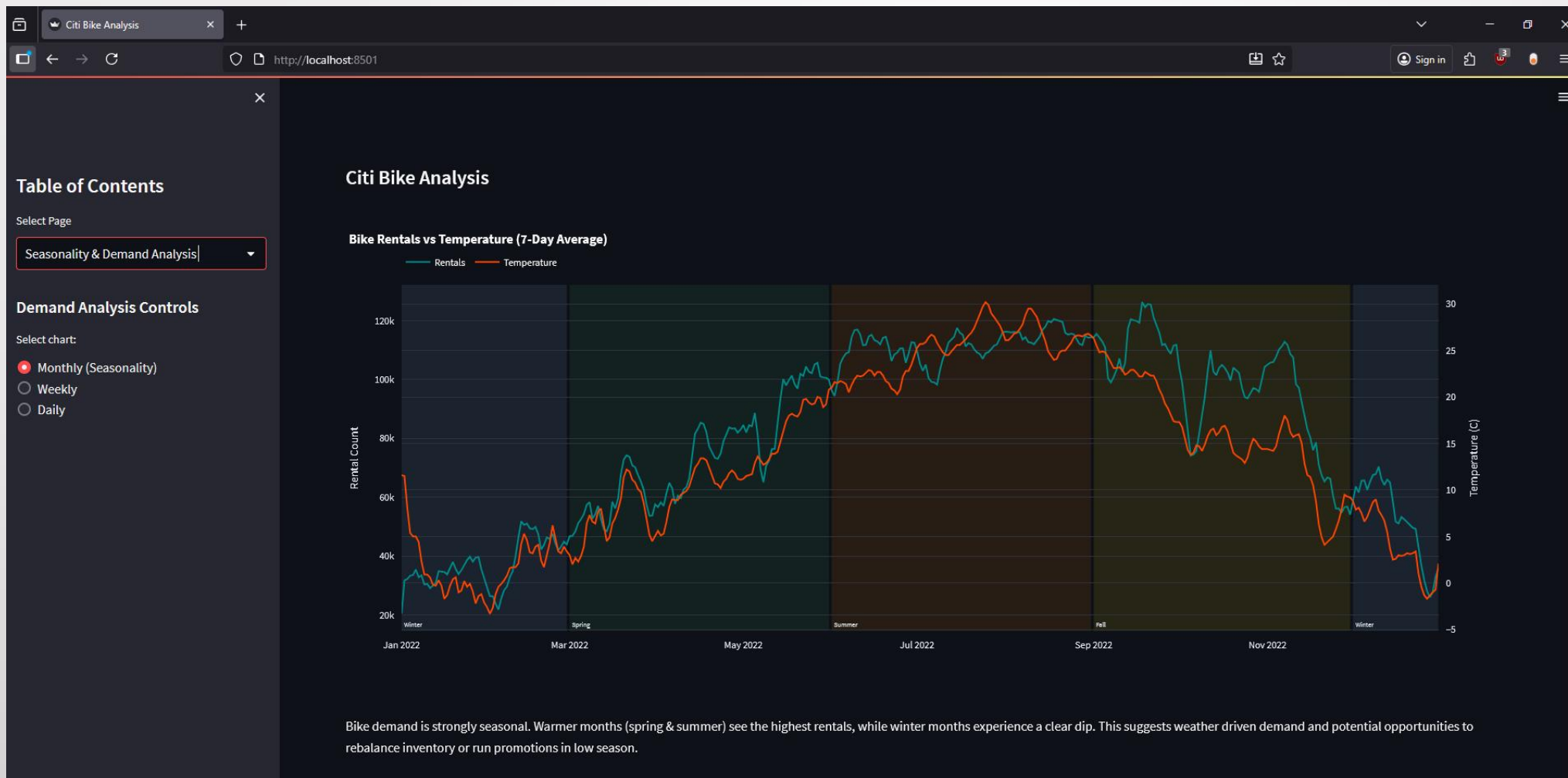


Key Questions

- What are some key factors driving demand?
- How can Citi Bike alleviate demand issues?







```
[6]: test = data.sample(frac = 0.01, random_state = 123)
```

Executive Summary & KPIs

+ 9 cells hidden

Seasonality & Demand Patterns

+ 19 cells hidden

Station Performance

Bar Chart - Top Stations Imbalance index - are some stations used more as start only/end only or more self sufficient (round trip) NYC has x number of stations

1 bar charts - Toggle: Top 20 start, top 20 end, Top start + end, Top Roundtrip 2 diverging bar charts - Imbalanced stations 1 bar charts - top balanced stations

```
[11]: station_df = data.loc[:,["start_station_name", "end_station_name"]]
```

```
[20]: top_station_df = \
(pd.concat([
    station_df["start_station_name"].value_counts().rename("departures"),
    station_df["end_station_name"].value_counts().rename("arrivals"),
    station_df[station_df["start_station_name"] == station_df["end_station_name"]]["start_station_name"].value_counts().rename("round_trips"),
    axis = 1
])
.fillna(0)
.assign(
    total_volume = lambda df: df["departures"] + df["arrivals"],
    imbalance_over = lambda df: df["departures"] - df["arrivals"], # stations that may be over supplied
    imbalance_under = lambda df: df["arrivals"] - df["departures"],
    imbalance_over_ratio = lambda df: np.where(df["total_volume"] > 100, # filter out stations with very low traffic & avoid dividing by 0
        df["imbalance_over"] / df["total_volume"],
        0),
    imbalance_under_ratio = lambda df: np.where(df["total_volume"] > 100, # filter out stations with very low traffic & avoid dividing by 0,
        df["imbalance_under"] / df["total_volume"],
        0),
)
.reset_index(names = "station")
)
```

```

1  # st_dashboard_2.py
2
3  C:\> Users > cneva > Projects > CitiBike Analysis > st_dashboard_2.py > ...
4
5  ##### Citi Bike Dashboard
6
7
8  ##### Imports Packages
9
10 import pandas as pd
11 import numpy as np
12 import matplotlib.pyplot as plt
13 import plotly.express as px
14 from plotly.subplots import make_subplots
15 import plotly.graph_objects as go
16 from datetime import datetime as dt
17 from keplergl import KeplerGL
18 import streamlit as st
19 from streamlit_keplergl import keplergl_static
20 from PIL import Image
21
22
23 ##### Import Data
24
25 top_station_df = pd.read_csv("Data/top_station_df.csv")
26 rentals_df = pd.read_csv("Data/rentals_df.csv")
27 KPI_df = pd.read_csv("Data/KPI_df.csv", index_col = "metric")
28 dow_df = pd.read_csv("Data/dow_df.csv")
29 hourly_df = pd.read_csv("Data/hourly_df.csv")
30
31
32
33
34 ##### Dashboard Set Up & Settings
35
36 st.set_page_config(page_title = "Citi Bike Analysis", layout = "wide")
37 # st.title("Citi Bike Analysis")
38
39
40 # Side bar
41 st.sidebar.title("Table of Contents")
42 page = st.sidebar.selectbox("Select Page",
43                             ["Executive Summary",
44                              "Station Performance",
45                              "Seasonality & Demand Analysis",
46                              "Interactive NYC Map",
47                              "Recommendations"])
48
49
50 ##### -----Pages
51 ##### -----Executive Summary
52 if page == "Executive Summary":
53     # st.markdown("### Executive Summary")
54     st.markdown("# Citi Bike Analysis Dashboard")
55
56     # read KPI_df
57     total_rentals = KPI_df.loc["rental count", "value"]
58     total_rentals_delta = KPI_df.loc["rental count", "delta"]
59
60     avg_duration = KPI_df.loc["average duration", "value"]

```



Marketing Strategy



Background

Instacart is an online grocery delivery and pickup service that connects customers with their own personal shopper who pick and pack groceries from local stores.



Objectives

Analyze Instacart's order history and customer base (32 million rows) to gain insights into retaining existing customers as well as acquiring new ones.



Key Skills

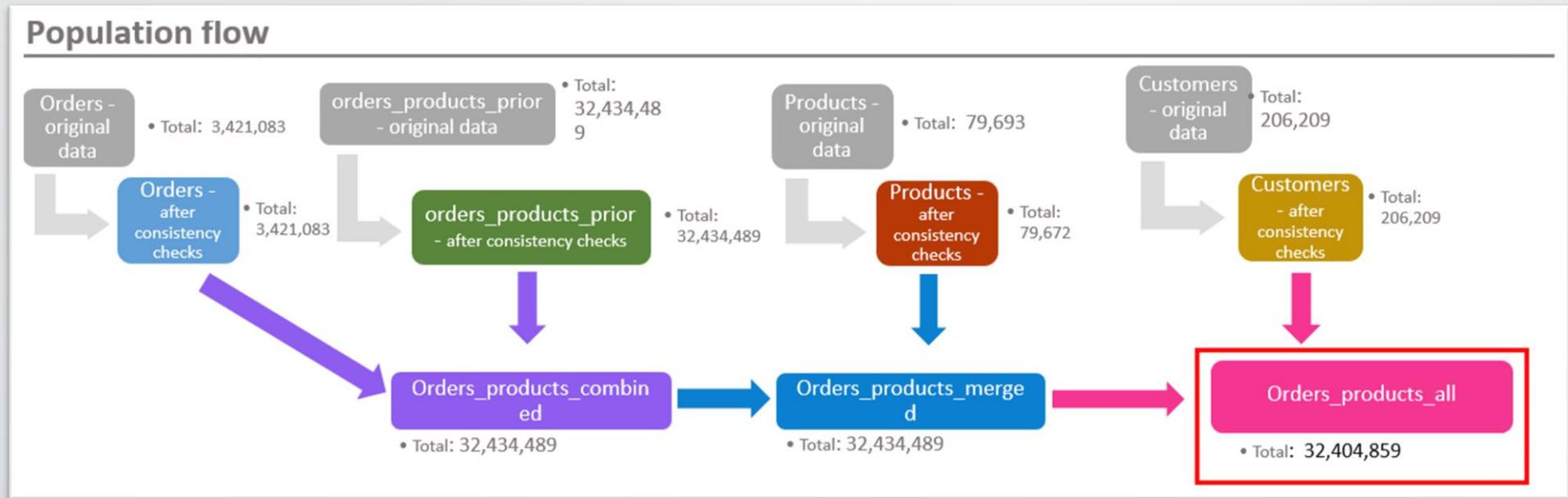
- Python – Pandas, NumPy, Matplotlib, Seaborn
- Data Cleaning & Data Wrangling
- Visualizations



Key Questions

- When are the busiest days of the week and hours of days?
- What products and departments drives sales?
- How can we group customers based on spending habits and customer profiling that drives marketing strategies?

Data Cleaning & Wrangling



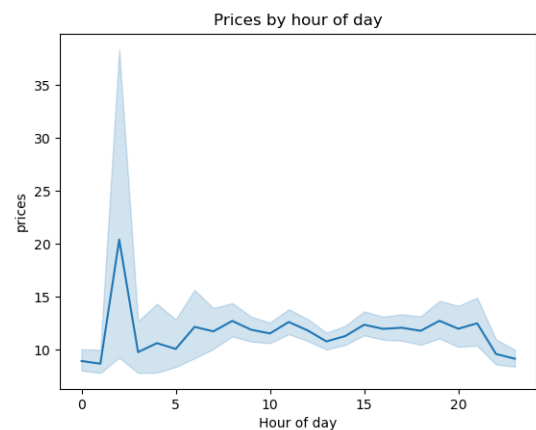
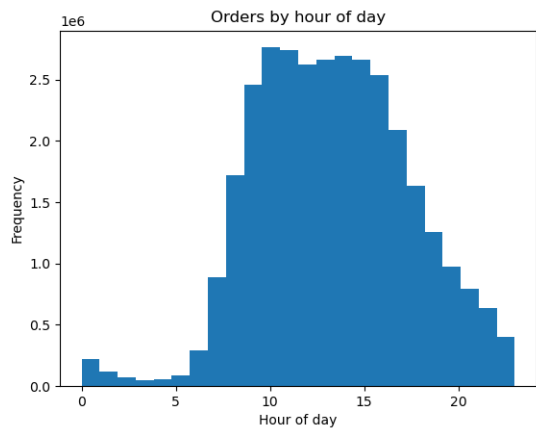
[Instacart Orders Dataset](#) is hosted on Kaggle
Customer Dataset was created by CareerFoundry for learning purposes

Pandas was used to clean, wrangle, and aggregate the data as well as create extra features to add dimensionality and allow customer segmentation all while ensuring accuracy

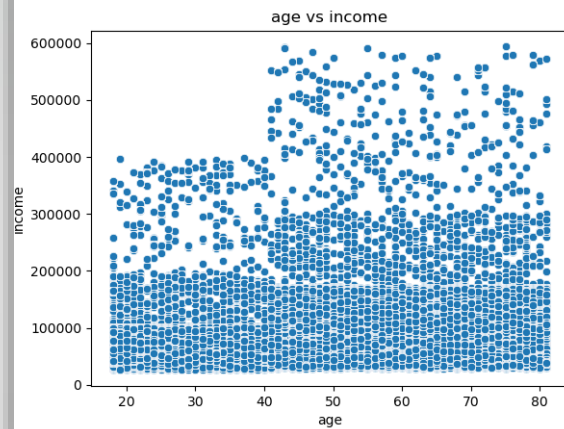
Coding best practices were followed to ensure no issues such as memory shortages when dealing with the large dataset (32 million rows)

Analysis

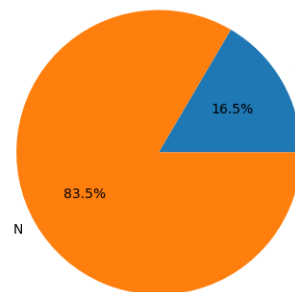
Order Trends



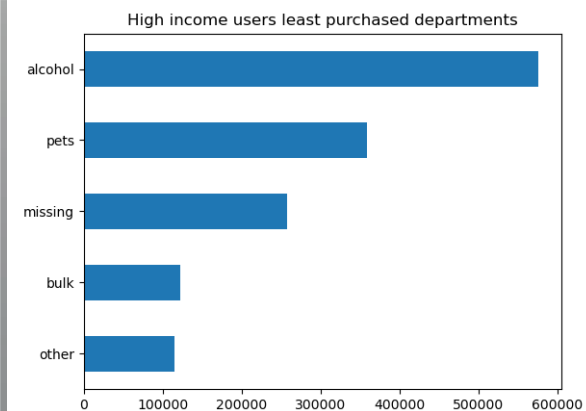
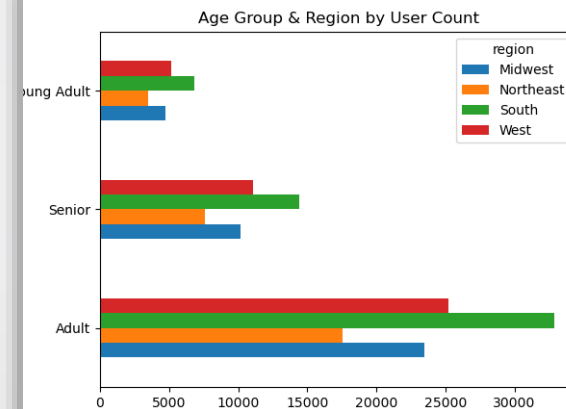
Customer Segmentation



Percent of Users That Are Single Adults



Marketing Targets



Insights & Recommendations



Maximize Efforts During Peak Hours

- Ensure capacity and marketing campaigns during peak hours where more traffic occurs
- Busiest ordering days are the weekends | Busiest hours are between 11AM – 2PM



Reward Repeat Customers

- Regular & Loyal customers make a majority of Instacart's customer base (75%+)
- Repeat customers are more likely to purchase from a wider variety of departments



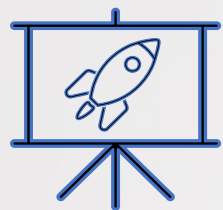
Incentivize New & Infrequent Customers

- Leverage limited promotions to cycle in new customers as well as refresh users who order infrequently
- New Customers have a higher total order value on average
- Give Infrequent customers (< 1 order per month) a greater reason to return to the service



Capitalize On High Value Customers

- Single & middle/high income adults is a large and growing customer segment | < 5% of products offered cost \$25+
- Introduce and incentivize high-range products to high value customers



Rockbuster Stealth

Launching a Video Rental Service



Background

Rockbuster Stealth LLC, is a fictitious video rental company trying to break into the stream service industry



Objectives

Analyze Rockbuster Stealth's database of movies and customers to create a strategy for transitioning to an online service model



Key Skills

- PostgreSQL – Querying a database using SQL
- Tableau – Visualization & Storytelling



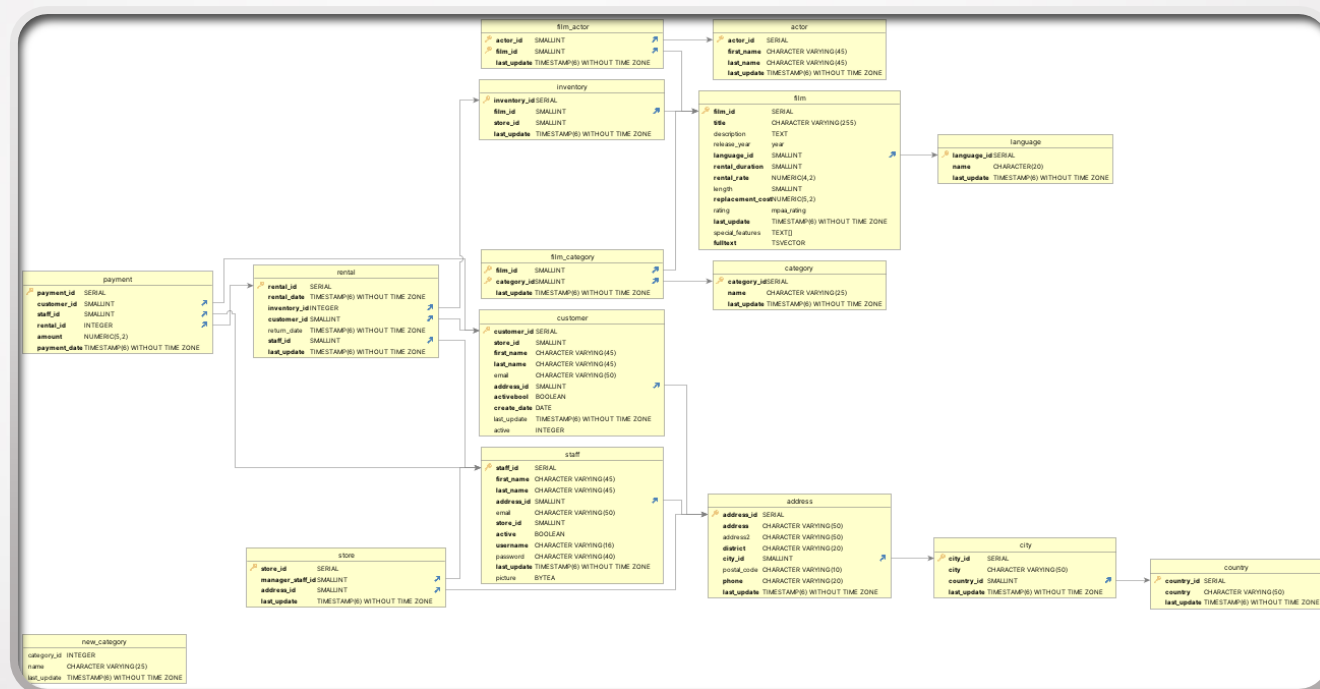
Key Questions

- What are the key movie characteristics that drive sales? Genre? Movie length? etc
- Which markets are the high value customers located in?

Database Querying

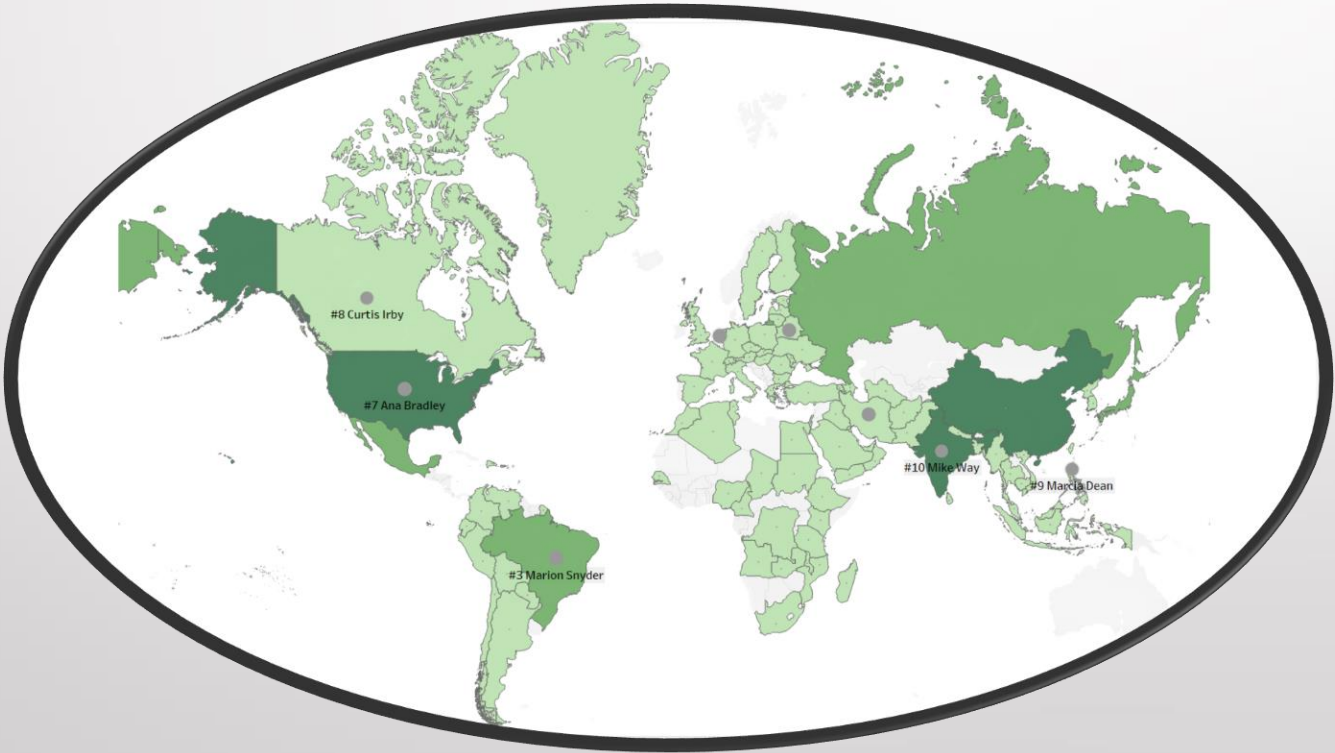
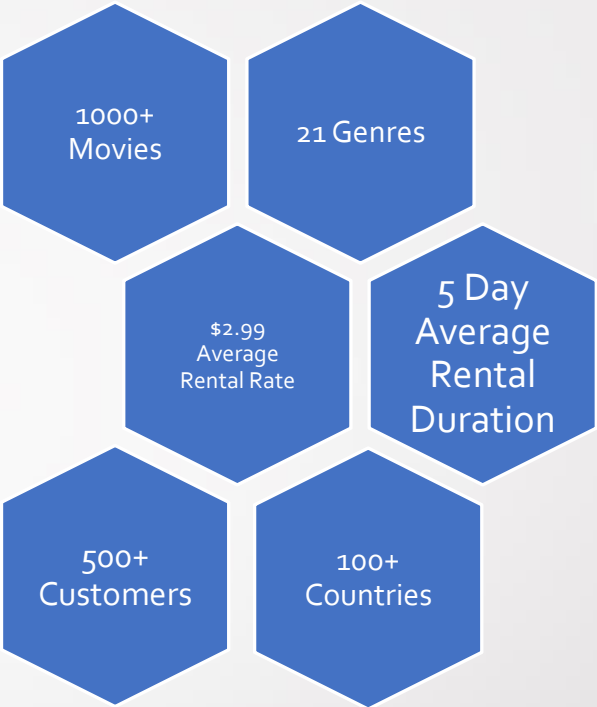
- Rockbuster Stealth's data is stored in a relational database of 16+ tables

- Extracting insights from the data involves efficiently querying the database using SQL



Descriptive Analysis

Sales at Rockbuster have been steadily decreasing therefore leadership has made the decision to pivot the business towards streaming as a service



Insights & Recommendations



Capitalize On Customer Preference

Rockbuster users enjoy a wide variety of genres.
Continue to invest in high performing genres while providing a diverse portfolio



Understanding Consumer Behavior

Optimize policy around consumer behavior
Consider promotions and specials around the average rental duration of 5 days



Expanding The Customer Base

Target marketing campaigns towards high-value markets, ie countries with high average customer value and low number of customers (China/India)



Pig E. Bank

Customer Risk Analysis



Background

Pig E. Bank is a fictitious global banking institution built to provide excellent and secure financial services



Objectives

Provide detailed analysis of Pig E. Bank's customers to improve customer retention and reduce security risks



Key Skills

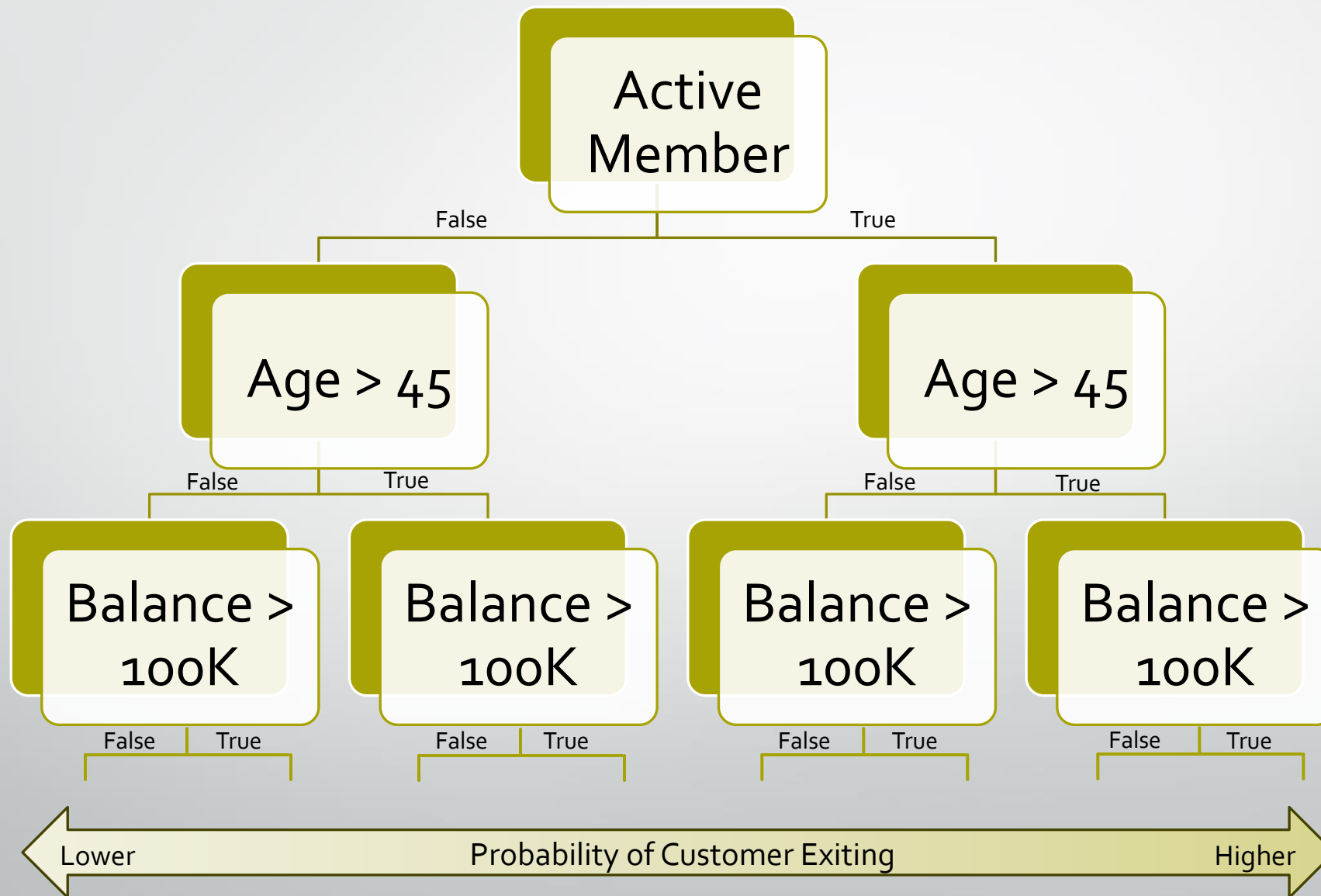
- Python – Pandas, Scikit-learn
- Machine Learning – decision tree algorithm & predictive analysis
- Time series forecasting



Key Questions

- What customer characteristics leads churn?
- How can we use predictive analysis to drive better decision-making?

Analysis – Decision Tree



Insights & Recommendations

Inactive Customers

- Customers with a low balance is more likely to close their account
- Incentivize customer activity through promotions

Age

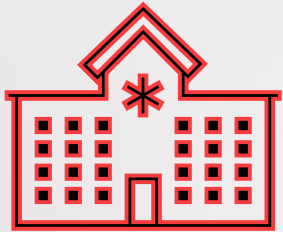
- Middle-aged customers are more likely to close their account
- Tailor customer care and products that this age group is more likely to engage in such as retirement planning

Product Portfolio

- Customers with only one financial product is more likely to close their account
- Ensure that new customers are fully aware of the bank's vast service offerings

Credit Score

- Customers with lower credit scores are more likely to close their account
- Identify early signs for customers with degrading credit scores



Preparing for Influenza Season



Background

In 2017, 45,000 people in the US have died from the flu and that number has been trending upwards. A medical staffing agency is preparing for the incoming influenza season



Objectives

The medical staffing agency would like data driven analysis on where and when to allocate their medical staff resources



Key Skills

- Tableau – Visualization & Storytelling with Dashboards
- Statistical Hypothesis Testing
- Correlation Analysis
- Forecasting

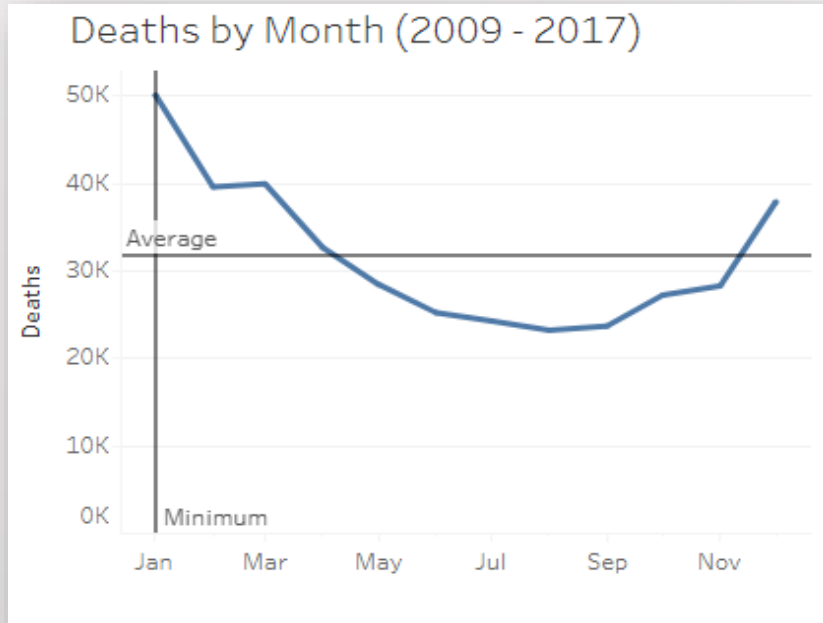


Key Questions

- What affects influenza season?
- What are the key demographics or vulnerable populations that may affect the influenza mortality rate more than others?



Identifying Key Contributors



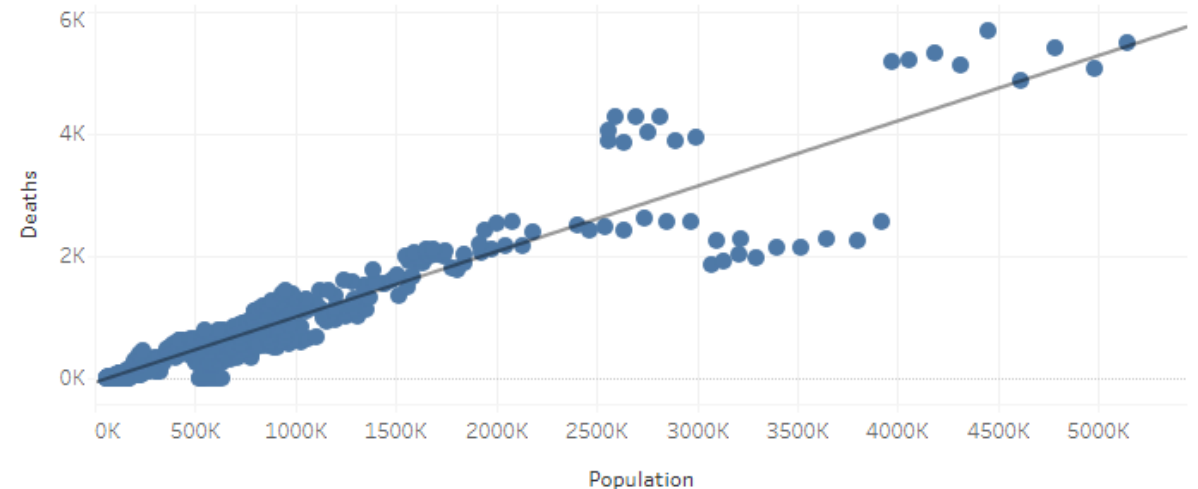
Age

- There is a strong positive correlation between age and influenza fatality rate
- States with a greater population, especially those with a greater relative elderly population would require more resources

Seasonality

- There is a strong negative correlation between temperature and influenza fatality rate
- States with a lower average temperature would require more resources

Correlation Between Population & Death - 65+ years



Insights & Recommendations



Population size

States with the most population should receive a proportional amount of resources



Age

States with a higher aging population should receive extra resources



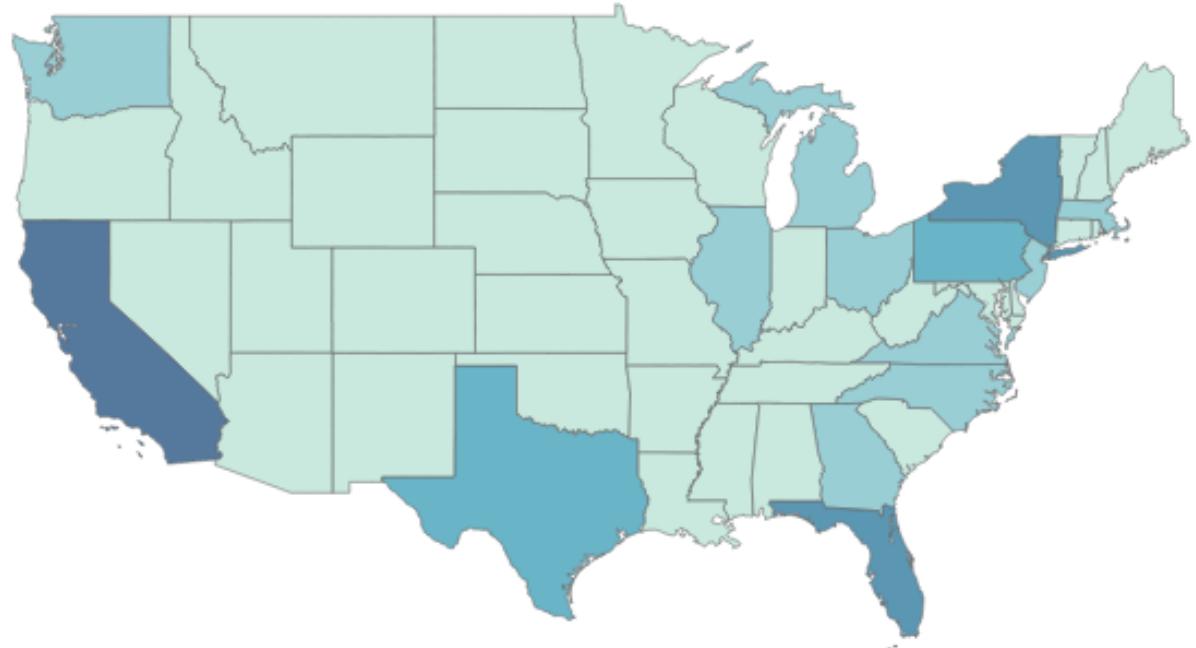
Temperature

States with a lower than average daily temperature should receive extra resources



Seasonality

The winter season should be the period of the year where the agency should look to increase recruiting



Weighted Allocation

0.2%



10.5%





GameCo Budget Analysis



Background

The video game customer base has shifted over the past decade. GameCo is looking make strategic business decisions that would recapture market share in a shrinking industry



Objectives

Analyze historical video game sales to understand the current customer behavior



Key Skills

- Excel
- Visualizing Results
- Descriptive Analysis



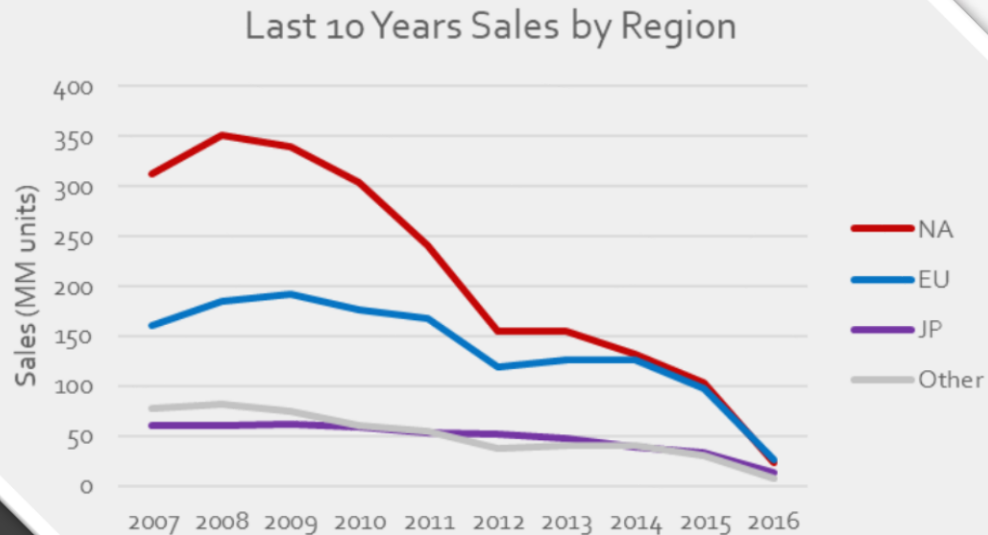
Key Questions

- What makes a popular video game? Genre? Platform? etc
- How can GameCo leverage its data to make better informed business decisions?

Analysis and Key Insights

Regional Allocation

- All regions have seen a decline in boxed video games over the last 10 years
- The NA region suffered the most severe decline



Platform Market

- The platform video games are enjoyed on shift depending on what is the current generation console
- The Playstation and Xbox continue to be the most dominant platforms

Sales (MM)		Year									
System	Platform	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Nintendo	3DS						125	102	112	86	54
Nintendo	DS	242	298	295	244	176	56	23	4	0	
Nintendo	GBA	11	7								
Nintendo	GC	23	1								
Nintendo	Wii	276	310	348	421	263	125	46	19	9	3
Nintendo	WiiU							36	44	45	33
Other	DC		0								7
Other	N/A	0	0	0	0	1					0
PC	PC	6	19	25	34	49	70	47	26	27	16
Sony	PS2	207	152	108	53	11	1				
Sony	PS3	42	148	239	265	289	319	219	235	102	36
Sony	PS4								50	198	230
Sony	PSP	112	95	69	76	70	36	15	6	0	0
Sony	PSV						10	37	25	29	14
Microsoft	X360	104	192	271	242	342	290	202	179	73	26
Microsoft	XB	20	1	0							
Microsoft	XOne								37	105	115

Analysis and Key Insights

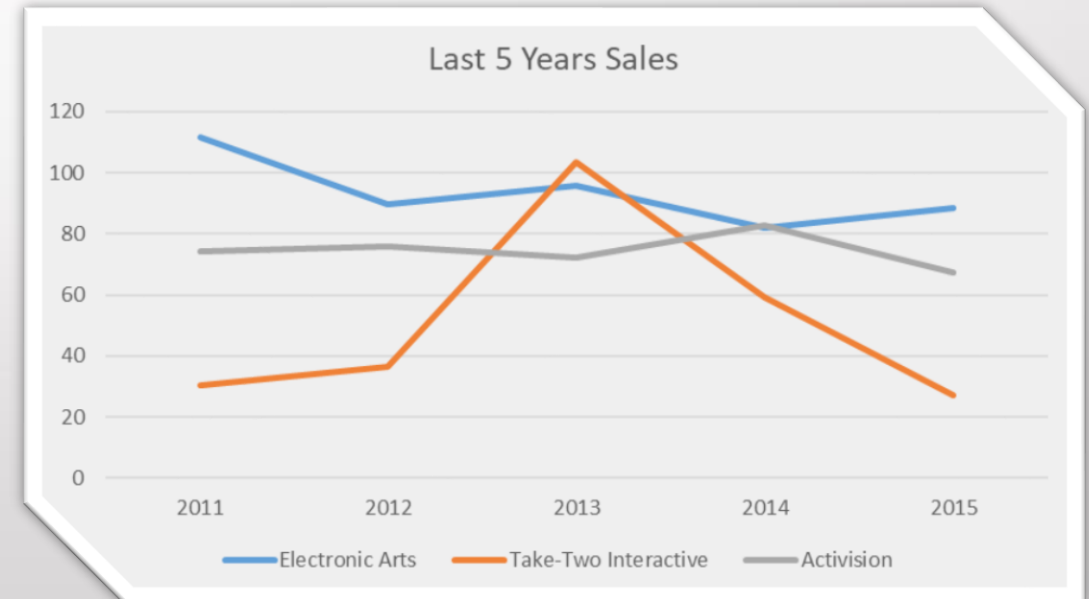
Focusing on Key Genres

- Action games have a huge volume of customers (28.1% share)
- Shooter games sell more per game on average (1.36 mil vs 0.80 market avg) leading to a potentially undersaturated market
- Role-Playing games are only one of two genres who has seen 3 years of continuous year over year growth

Competition Analysis

- Electronic Art dominates the Sports genre, therefore not a recommended genre to pursue
- Take-Two Interactive and Activision are both on the decline in recent years for any competing titles GameCo considers creating

Genre	Last 10 Year Sales	Avg Sales per Game	Sales versus Prior Year
Action	749.8	0.59	
Adventure	61.5	0.12	
Fighting	123.3	0.46	
Misc	194.0	0.44	
N/A	0.2	0.20	
Platform	67.3	0.57	
Puzzle	3.0	0.10	
Racing	173.9	0.61	
Role-Playing	235.7	0.46	
Shooter	594.0	1.36	
Simulation	32.5	0.31	
Sports	405.1	0.60	
Strategy	25.4	0.20	
Grand Total	2,665.5	0.56	



Key Takeaways

The boxed video game industry is on a decline over the past 10 years. In order to maintain and increase market share against the competition, GameCo. should take initiative of the following:

- Shift marketing resources from NA to EU
- Develop games on current/new generation of platforms, ie. PS4/Xone
- Focus efforts on critical genres, ie. Action, Shooter, Role-Playing