JEREMY OBACH

Data Analyst Portfolio

Project List

ANALYZING GLOBAL VIDEOGAME SALES

FLU SEASON STAFFING IN THE US

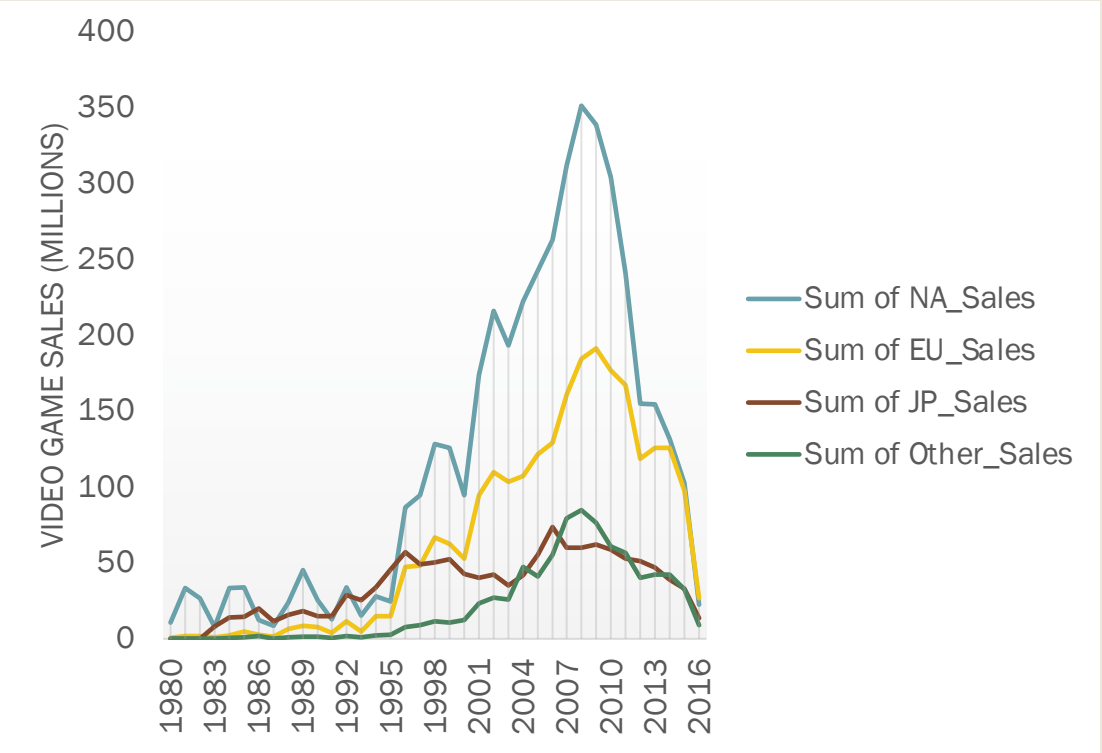
ROCKBUSTER VIDEO RENTAL DATABASE

TARGETED MARKETING STRATEGY WITH INSTACART

AUSTIN, TX AIRBNB MARKET ANALYSIS

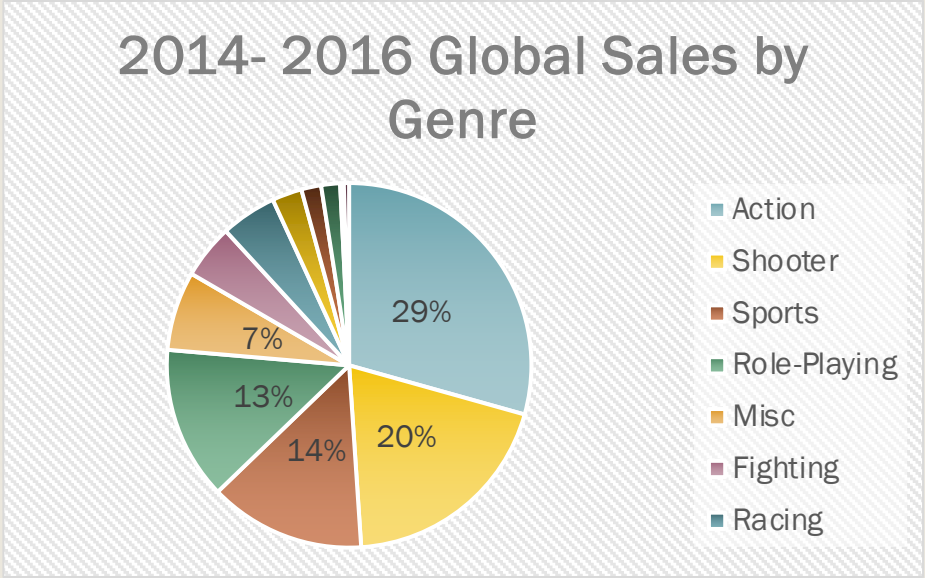
Analyzing Global video game sales

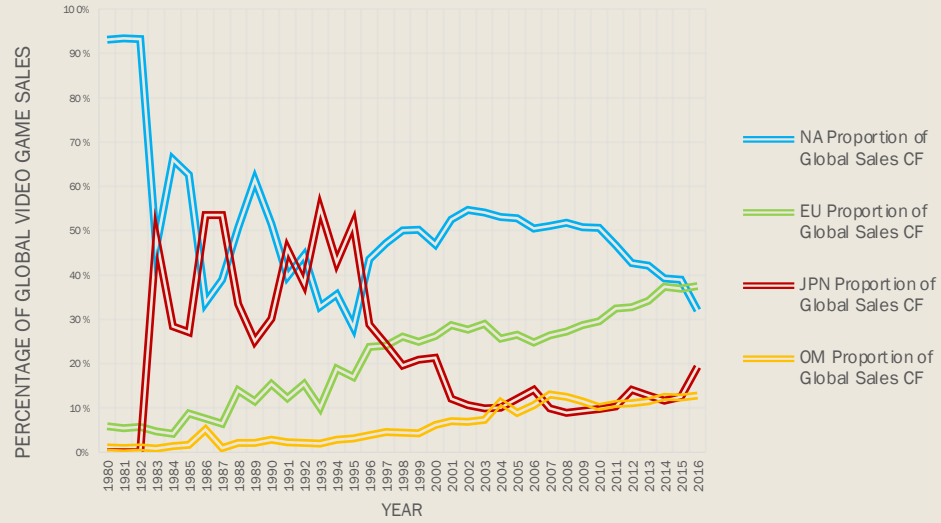
- Task: Provide GameCo executives with insights on the sales data to determine what type of game they should develop.
  - *Dataset: Video game titles released from 1980 to 2016 including genre, console, and regional market sales figures.*



Analyzing Global video game sales

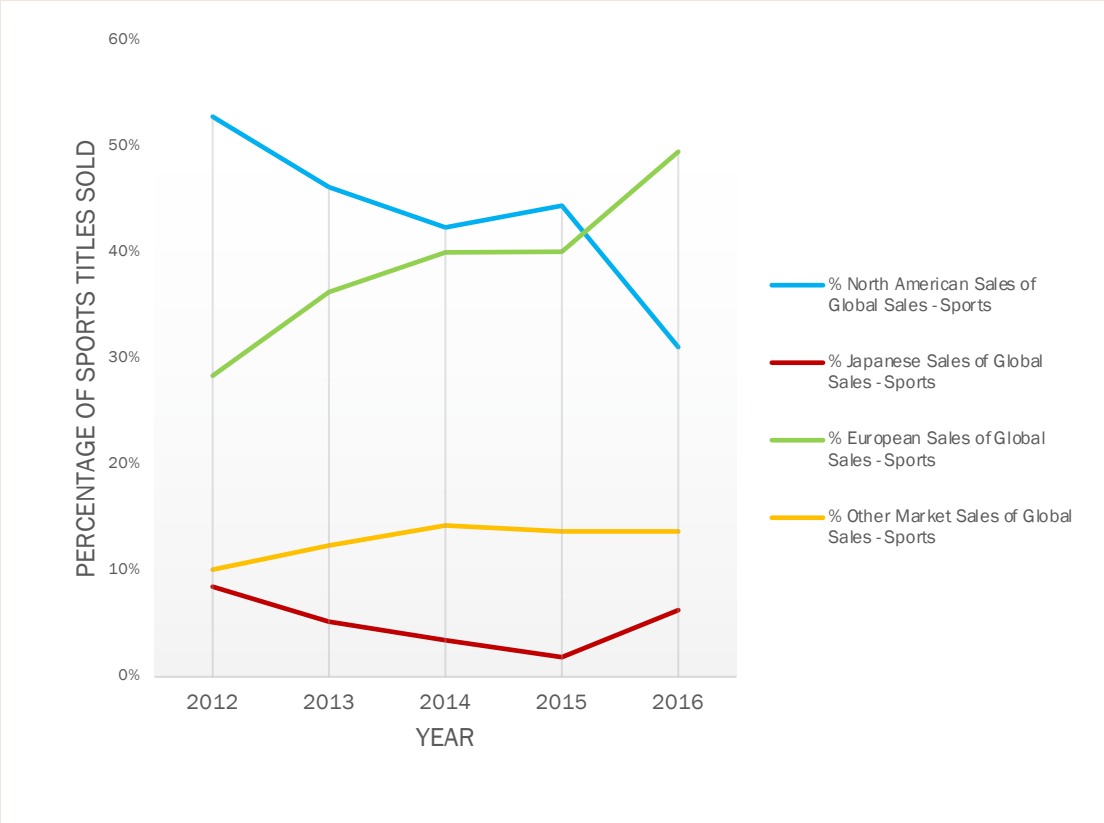
- Analysis – used Excel's pivot table and chart functions to identify trends in sales data.
  - *Top 4 genres account for 76% of recent sales*
  - *European market taken narrow lead on US in 2016.*
- Isolating precipitous drop in physical sales (digital sales not accounted for in dataset), we examined the last 3 years in sales by genre.

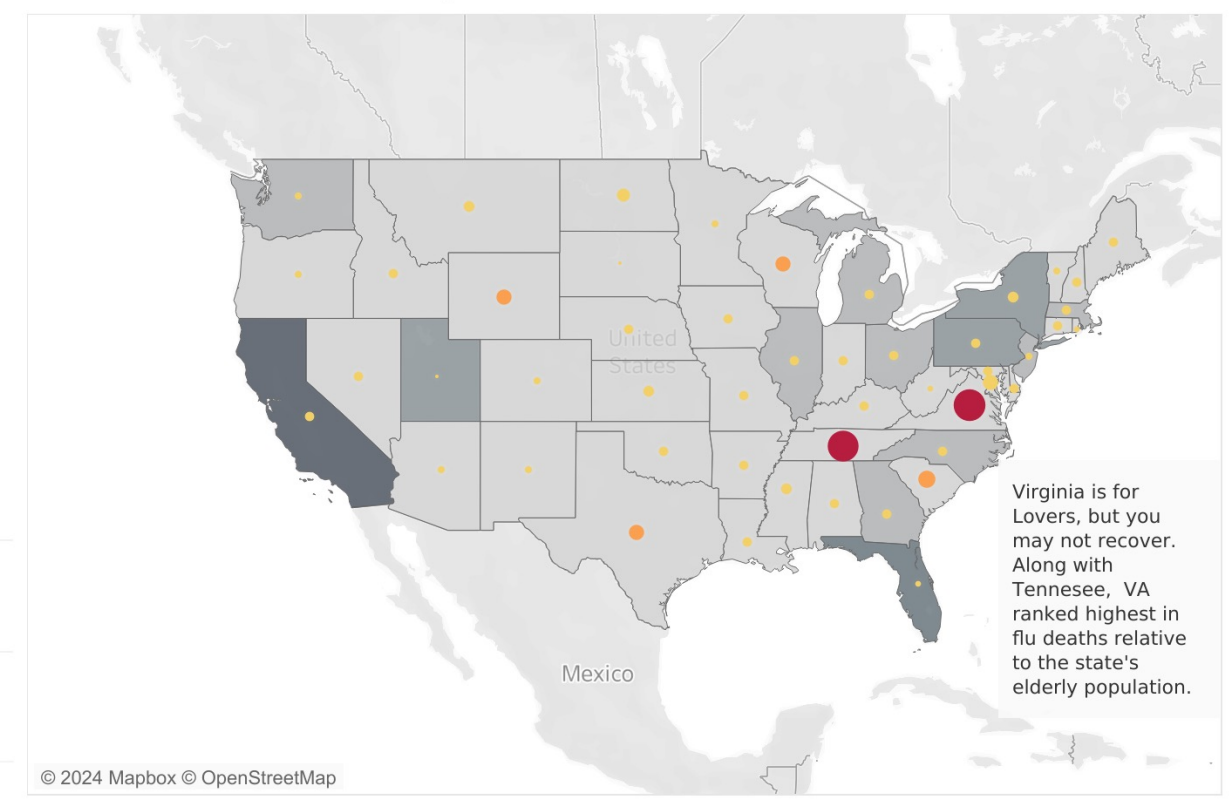




Analyzing Global video game sales

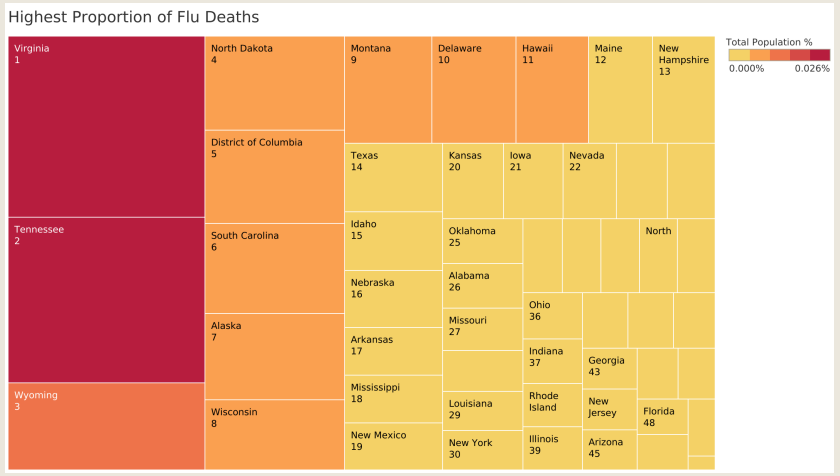
- Recommendation: develop sports title, popular across markets but especially in growing EU, and relatively unsaturated compared to other top genres.
  - *Release on PS4, XBOX One S, and PC.*
- Tools Used:
  - *Excel*
  - *PowerPoint*

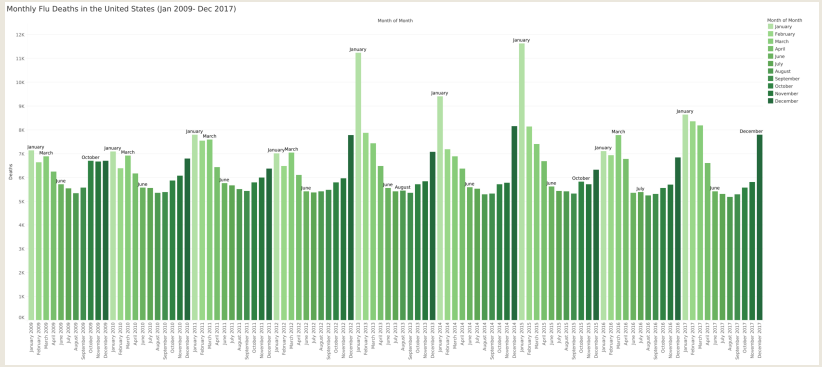




Flu season staffing in the US

- Task: Advise medical staffing company on where and when to send additional staff to most effectively curb influenza deaths
  - *Datasets:*
    - CDC Influenza-related deaths
    - US Census Bureau population data





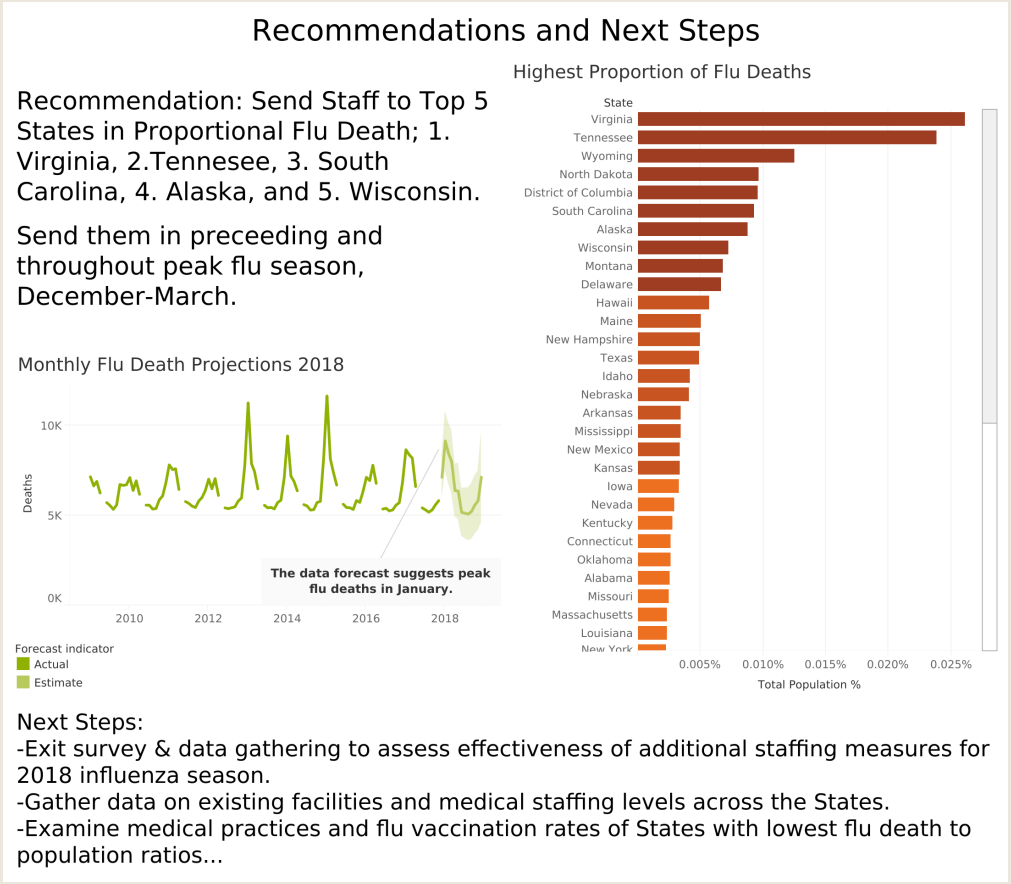
Flu season staffing in the US

- Analysis - Derived States with highest ***proportion*** of flu deaths to total population, to see which states are underperforming given their size.
  - *Though the most flu deaths were invariably in the States with highest population count, these states also had the highest medical staff numbers.*
- Next identified seasonality in flu deaths, reliably peaking in January.

Flu season staffing in the US

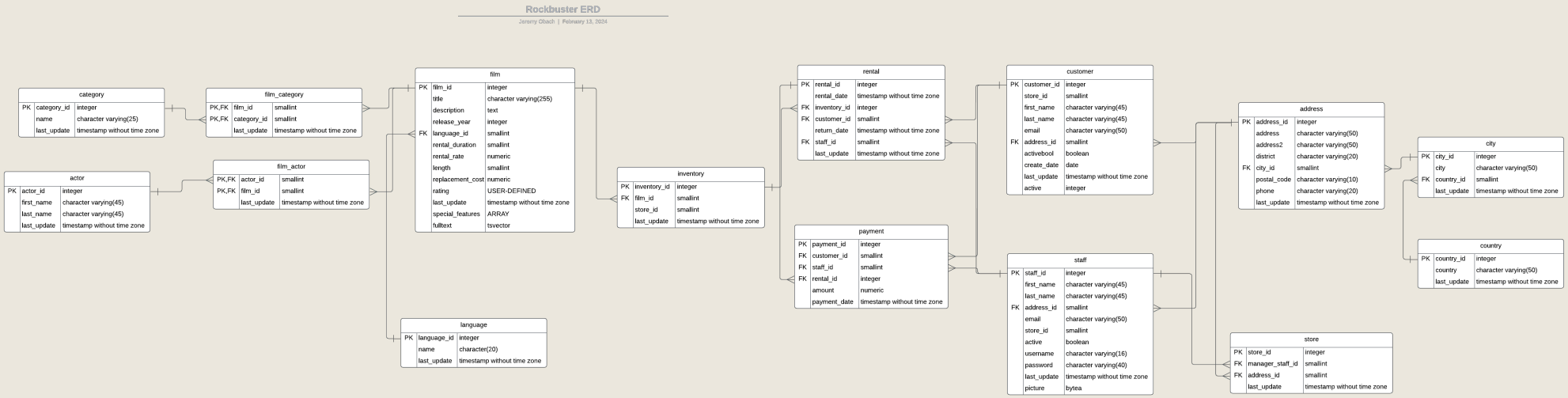
- Recommendation – send staff to top 5 states in proportional flu deaths, in the months preceding and during peak flu season
  - *VA, TN, WY, ND, SC.*
  - *December thru March*

- Tools Used + Links
  - *Tableau*
  - *Loom*
  - *Excel*

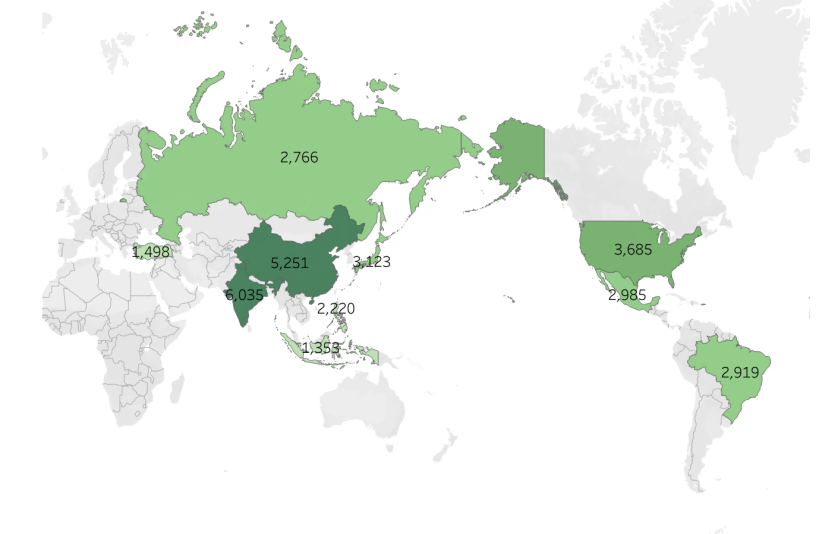


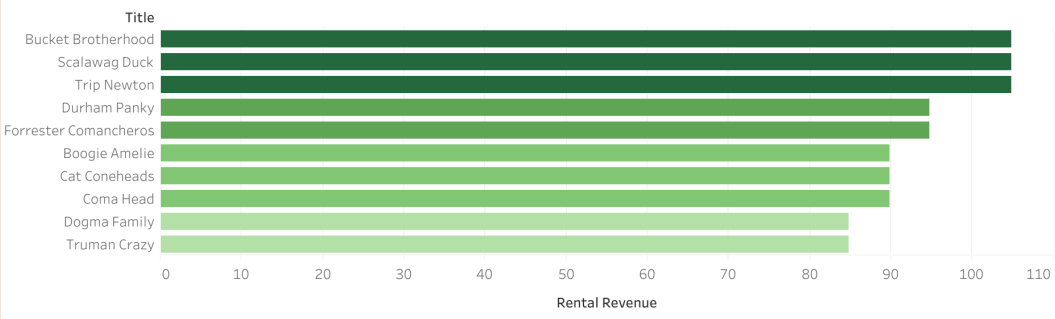
ROCKBUSTER VIDEO RENTAL DATABASE

- Task – help Rockbuster executives understand their customer-base and top performing films.
  - Create ERD and data dictionary for future employees navigating Rockbuster's RDBMS
  - *Dataset: relational database accessed thru PostgreSQL for 'Rockbuster', a movie rental company.*

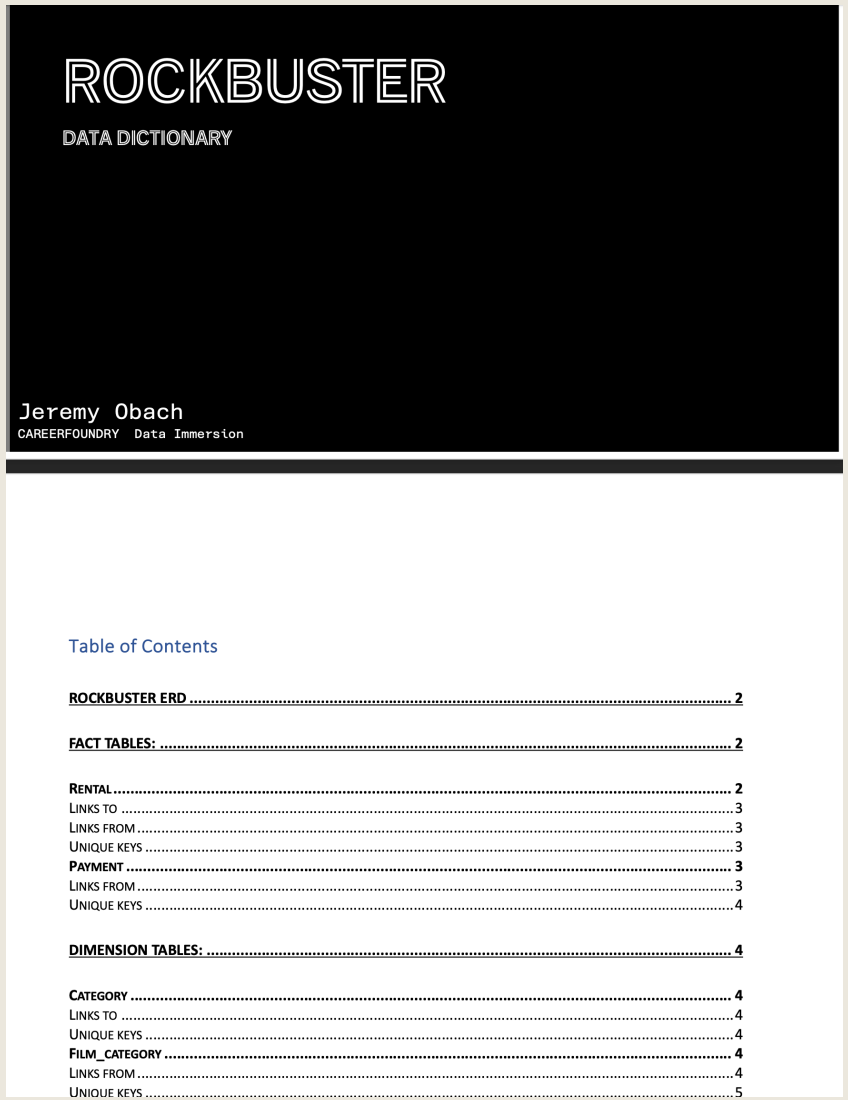


ROCKBUSTER VIDEO RENTAL DATABASE



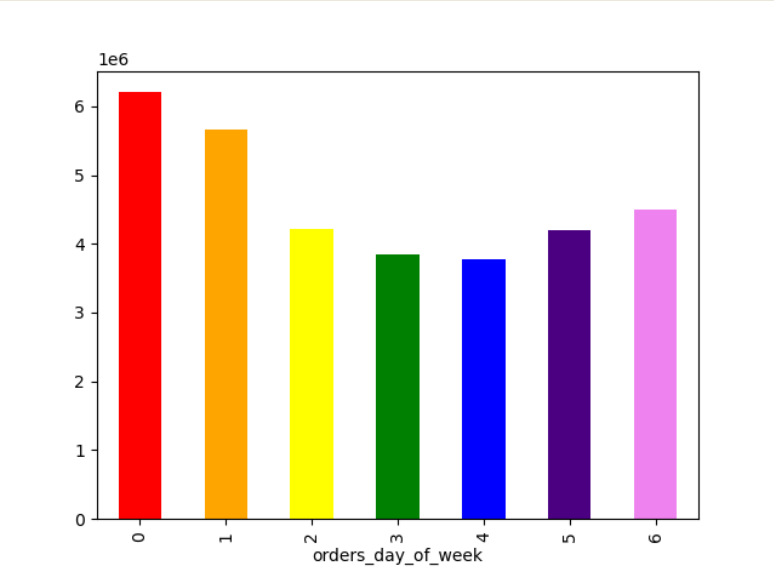


- Analysis
  - Identified top 10 countries, top 10 cities, and top 5 customers by rental revenue.
  - Identified top 10 performing titles by revenue.
  - Compared CTE vs Subquery performance
  - Derived descriptive database statistics
    - *Mean, median, mode*
  - Created ERD and Data Dictionary for Rockbuster database



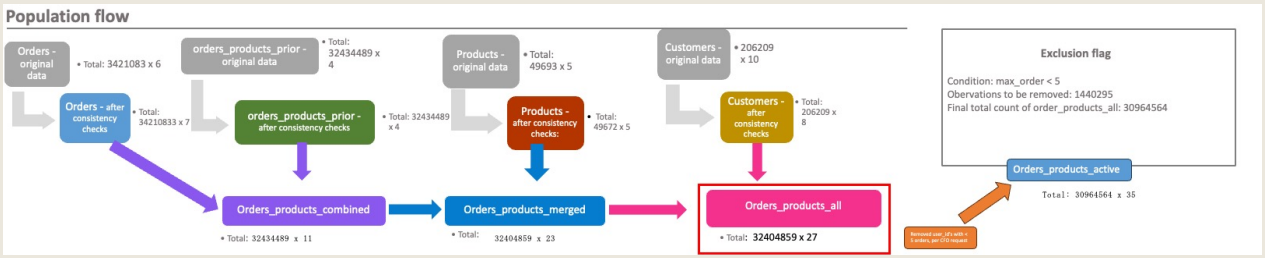
ROCKBUSTER VIDEO RENTAL DATABASE

- Key Insights
  - Average rental duration ~5 days
  - Average run time ~115 minutes
  - Wide distribution of customers
    - *314 cities, only 1 city with more than 1 customer*
    - *2 of top 5 customers in Aurora, CO*
- Tools Used + Links:
  - *PostgreSQL*
  - *GitHub*
  - *Excel*
  - *Word (data dictionary)*
  - *Tableau*
  - *Lucid*
  - *PowerPoint*



Targeted Marketing Strategy with Instacart

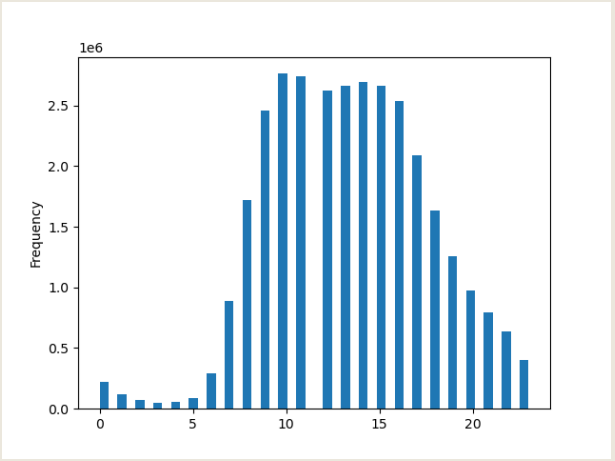
- Task – Provide Instacart marketing team with insights on:
  - Busiest hours of day, days of week
  - Price distribution on items ordered
  - Order frequency/ popularity by department
  - Customer profiles
    - *Dataset: open-sourced Instacart product, customer, and order data.*
      - *Several datasets; needed cleaning and merging. See Population flow*



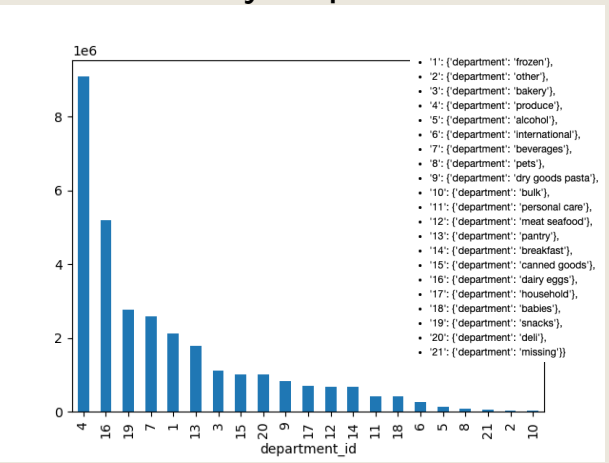
Targeted Marketing Strategy with Instacart

- Analysis- derived variables such as:
  - loyalty flags to identify frequent customers
  - LTV to get sum of product prices per user_id
  - Region and Generation tags based on customer State and Age, respectively
  - Customer profiles based on dependents, familial status, age, etc.

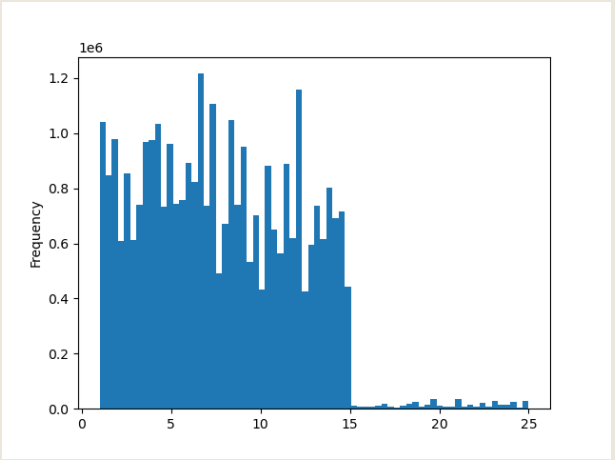
Order times



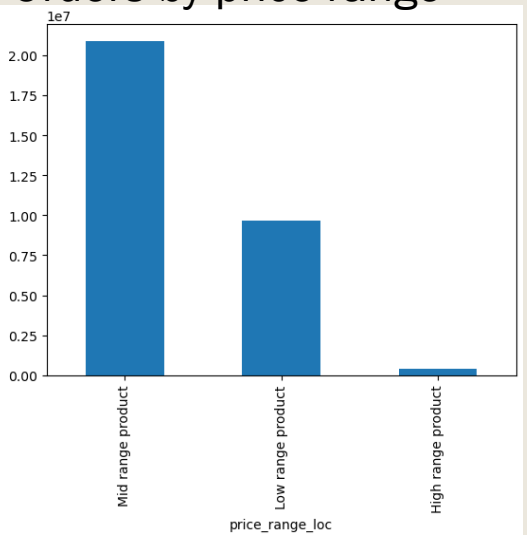
Orders by department



Item Price distribution



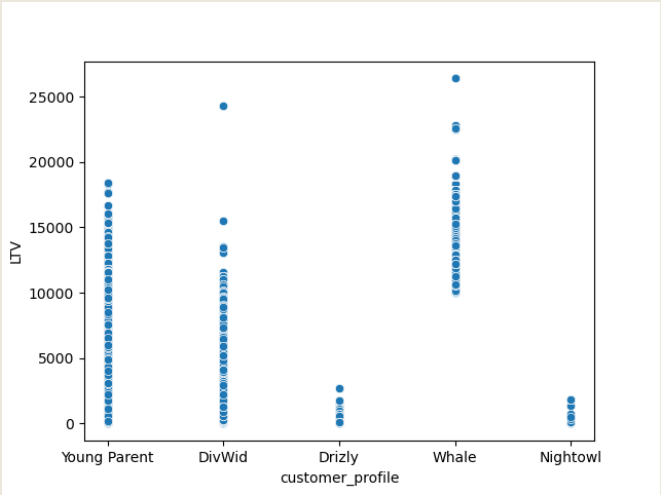
Orders by price range



Targeted Marketing Strategy with Instacart

- Recommendations
  - Advertising to bolster orders at non-peak days and hours, particularly weekday evenings
  - Increased advertising and selection for top 5 department categories
  - Develop targeted advertising, especially toward Divorced/Widows and young parents.

Python visualization using Seaborn



Aggregate the min/mean/max on a customer-profile level for usage frequency (orders) and expenditure (amount spent)

```
df.groupby('customer_profile').agg({'max_order':['mean','min','max']})
```

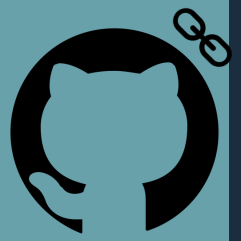
customer_profile	max_order mean	max_order min	max_order max
DivWid	34.348694	5	99
Drizly	15.355394	5	74
Nightowl	9.059677	5	18
Whale	77.985122	27	99
Young Parent	35.443911	5	99

```
df.groupby('customer_profile').agg({'LTV':['mean','min','max']})
```

customer_profile	LTV mean	LTV min	LTV max
DivWid	3315.963933	15.7	24307.1
Drizly	446.330930	7.7	2723.2
Nightowl	630.360484	45.2	1815.7
Whale	12630.166451	10014.1	26394.9
Young Parent	3566.815369	13.3	18448.1

Excerpt from Jupyter notebook

Tools Used + Links



jupyter 4.9.2 Task - visualizations with ords_prods_all Last Checkpoint: 01/16/2024 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help Not Trusted Python 3 (ipykernel)

Run Markdown

Exercise 4.9 Part 2

Contents

- Importing analysis and visualization libraries
- Importing ords_prods_all df

3. Created a histogram of the "orders_hour_of_day" column
4. Bar Chart demonstrating distribution of orders among customer loyalty designation
5. Line chart demonstrating avg item expenditure based on time of day
   - utilized sampling method
6. Line chart demonstrating connection between age and number of dependants (or lack thereof)
7. Scatterplot demonstrating link between age and spending power (income)

- trimming off unnecessary columns

8. Exporting df via pickle

Importing analysis and visualization libraries

In [3]:
```
# Import libraries

import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
import seaborn as sns
import scipy
```

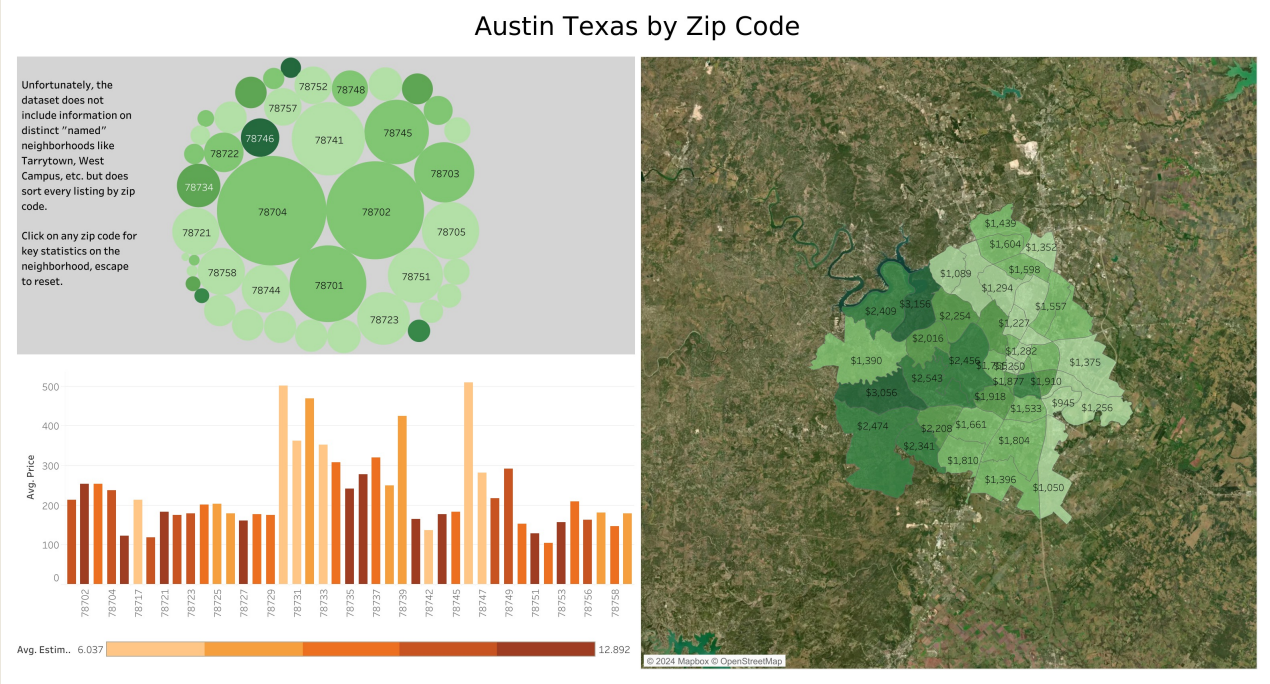
In [4]:
```
# Setting path to Achievement 4 master folder
path = r'/Users/jeremyobach/Documents/Data Analytics/CareerFoundry/Achievement 4- Python for DA/DEC23 Instacart Bask
#checking path set correctly
path
```

Out[4]:
```
'/Users/jeremyobach/Documents/Data Analytics/CareerFoundry/Achievement 4- Python for DA/DEC23 Instacart Basket Anal
ysis - MASTER FOLDER'
```

Excerpt of Jupyter notebook

Austin, TX Airbnb Market Analysis

- Task – evaluate Austin, Texas short term rental market for investment viability.
  - *Determine patterns, trends, opportunities in the ATX STR (short term rental) market.*
- Dataset: web scrape of Austin, TX Airbnb listings and reviews data

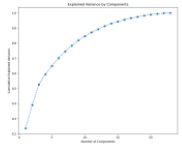


Austin, TX Airbnb Market Analysis

- Analysis
  - *Used machine learning techniques, geospatial and time series analysis to uncover insights as to the current state of the Austin, TX short term rental (STR) market.*
    - Clustering, an unsupervised machine learning technique, yielded statistical insights on successful vs unsuccessful listings.

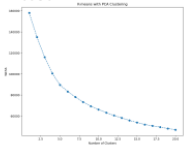
Clustering: Unsupervised Machine Learning Technique

Using **Scikit-learn**, a machine learning library for Python, I analyzed the listings dataset and determined that 80% of its variance came down to 9 components.

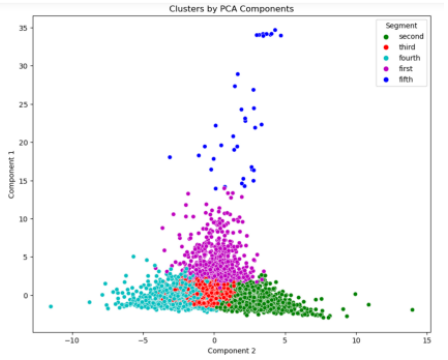


From there, I standardized the data using StandardScaler(), as the different components (pricing, review scores, number of reviews, age of listing in months) all had different orders of magnitude.

Then I applied PCA (Principal Component Analysis) followed by a k-means clustering algorithm in order to reduce the datasets dimensions and determined the number of clusters, which was 5 segments, via the elbow test.



After plotting a number of variable pairs with their clusters, and aggregating descriptive statistics for key variables to distinguish the segments, I was able to derive some key takeaways that distinguish the segments from one another. ->




```
df_pca_kmeans['Segment'].value_counts()

Segment
third     4357
second    1989
fourth    1287
first      733
fifth       38
Name: count, dtype: int64
```

Takeaways from Descriptive Statistics:

first segment - new and mediocre listings

- 2nd lowest listing age statistics (in months), only ahead of fifth segment.
- 2nd lowest median estimated revenue, 2nd lowest review scores.

second segment - established, low margin, high volume.

- highest mean number of reviews by a fair margin, and yet only 3rd in review numbers by median (3rd and fifth categories have wide margin between median and mean)
- lowest average prices by mean, and 2nd lowest (to first segment) in median.
- High review scores, but still 3rd behind fourth and third segment.
- Second highest estimated revenue, behind fourth segment, in both mean and median.

third segment - middle of everything, checks the box.

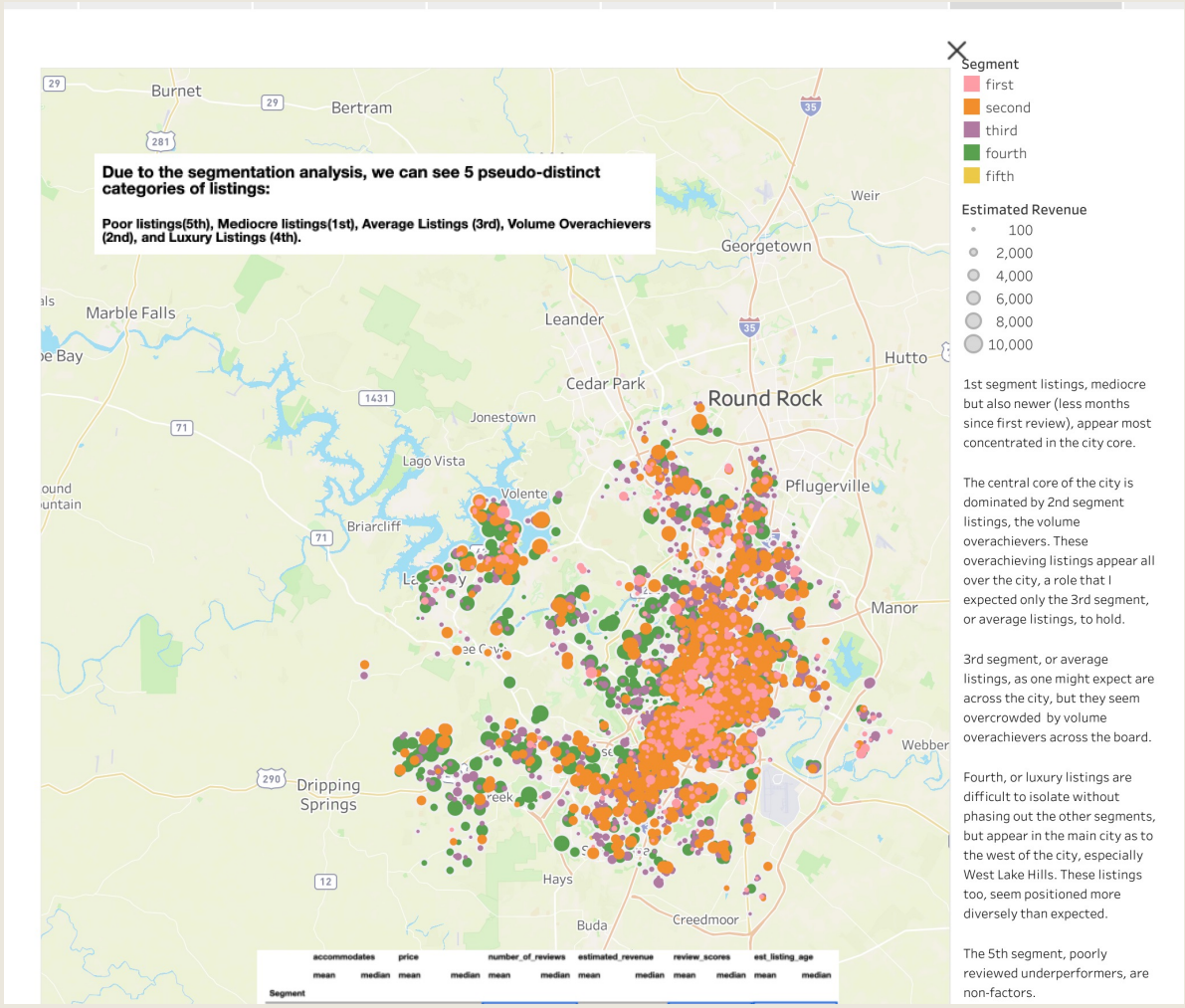
- middle in price by mean
- middle by review count
- middle estimated revenue
- curiously, highest review scores
- Also tied for highest median listing age, close first for mean listing age.

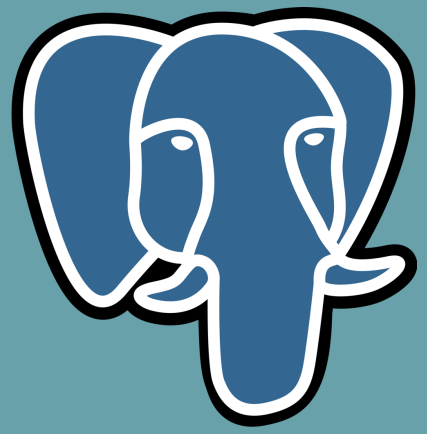
fourth segment - luxury/upscale/ large listings

- highest accomodation numbers by a wide margin: 10+ vs 4-5 for the other segments
- Highest price by a wide margin: median 407 vs other segments in the 100s.
- 2nd highest reviews, but not close to second segment.

Austin, TX Airbnb Market Analysis

- Recommendations:
  - *Combining clustering and geospatial analysis revealed likely saturation across Austin neighborhoods.*
    - to stand out a listing must be priced competitively with high turnover and positive reviews or be a large/ luxury listing.
- Tools Used:
  - *Anaconda/ Python Libraries:*
    - Pandas
    - SciKit Learn
    - Matplotlib
    - Seaborn
  - *Tableau*












Skillset Recap

Contact information

- https://www.linkedin.com/in/jeremy-obach/
- https://github.com/jobachone/
- Jeremy.Obach@gmail.com

- Resume ->