



Rich Bradley

## Data Analytics Portfolio

<https://www.linkedin.com/in/rich-bradley-mba/>

## Table of Contents

[GameCo Market Analysis](#)

[Preparing for Influenza Flu Season](#)

[RockBuster Data Analysis](#)

[InstaCart Grocery Basket Analysis](#)

[Flight Delay Project Analysis](#)





# Project 1 - GameCo

## Data Set

- Total number of units of games sold between 1980 to 2016.
- Numbers represent units sold in millions.
- Regions North America, Japan, and Europe.
- Different platforms, genres, and publishers.
- Data Excel file.
- Source: VGChartz

## Key Questions

- Are certain games more popular than others?
- Any games increase/decrease in popularity?
- How have the sales figures varied between geographic regions over time?

## Tasks

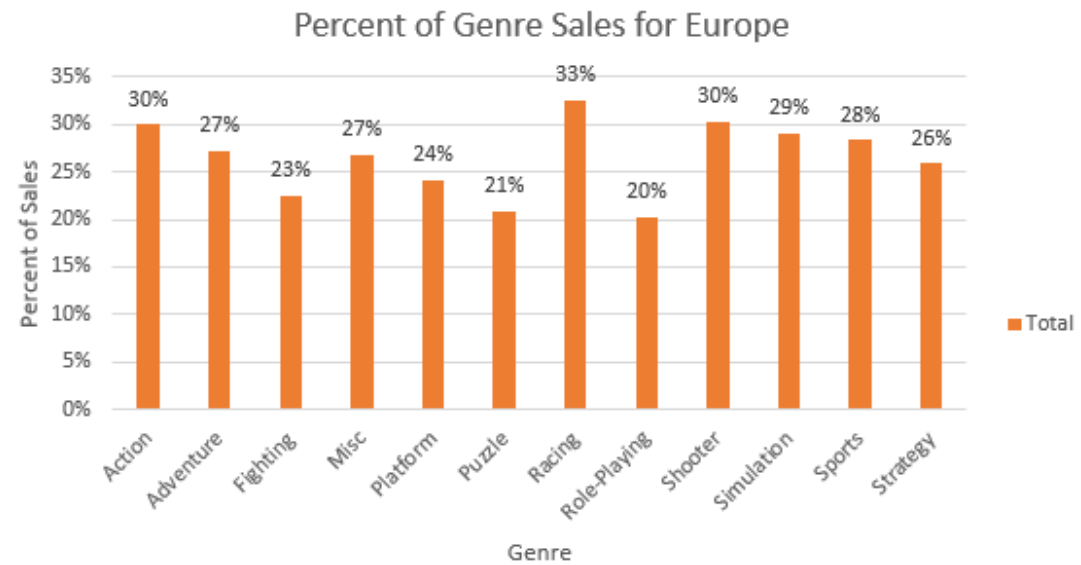
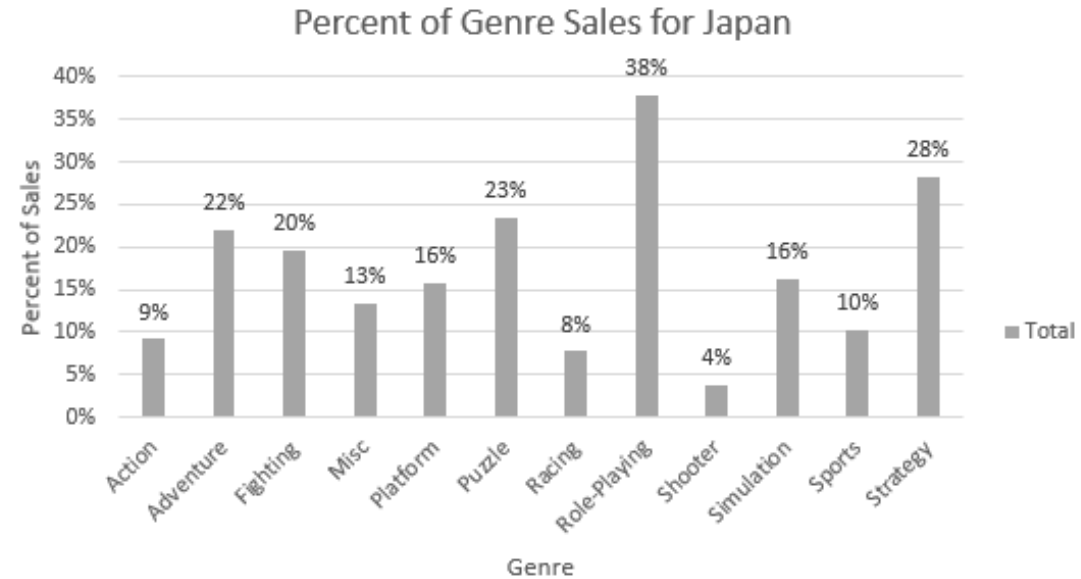
- Cleaning and formatting data.
- Build pivot tables.
- Grouping the data
- Conduct descriptive analysis and develop visual insights



# GameCo

Role Playing games are the highest sales at 38% in Japan, while Shooter games are the least with 4% in sales..

Racing games account for 33% in sales for Europe and Role-Playing games are 20% of sales.



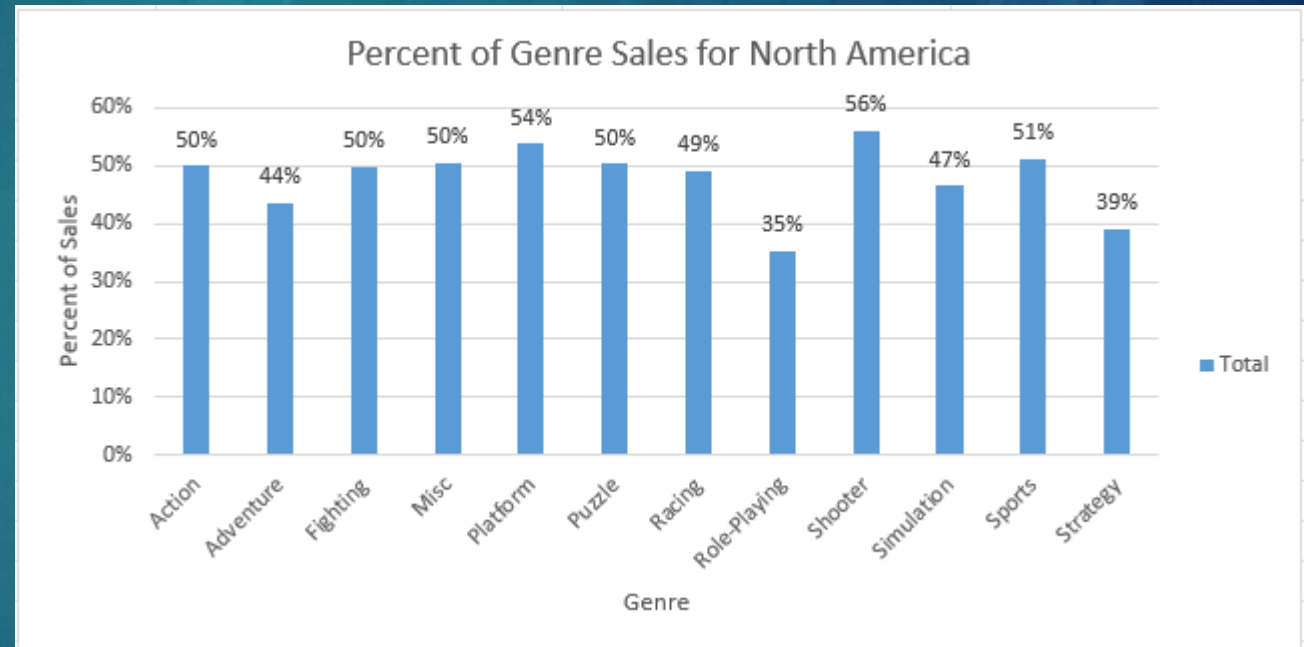




# GameCo

5

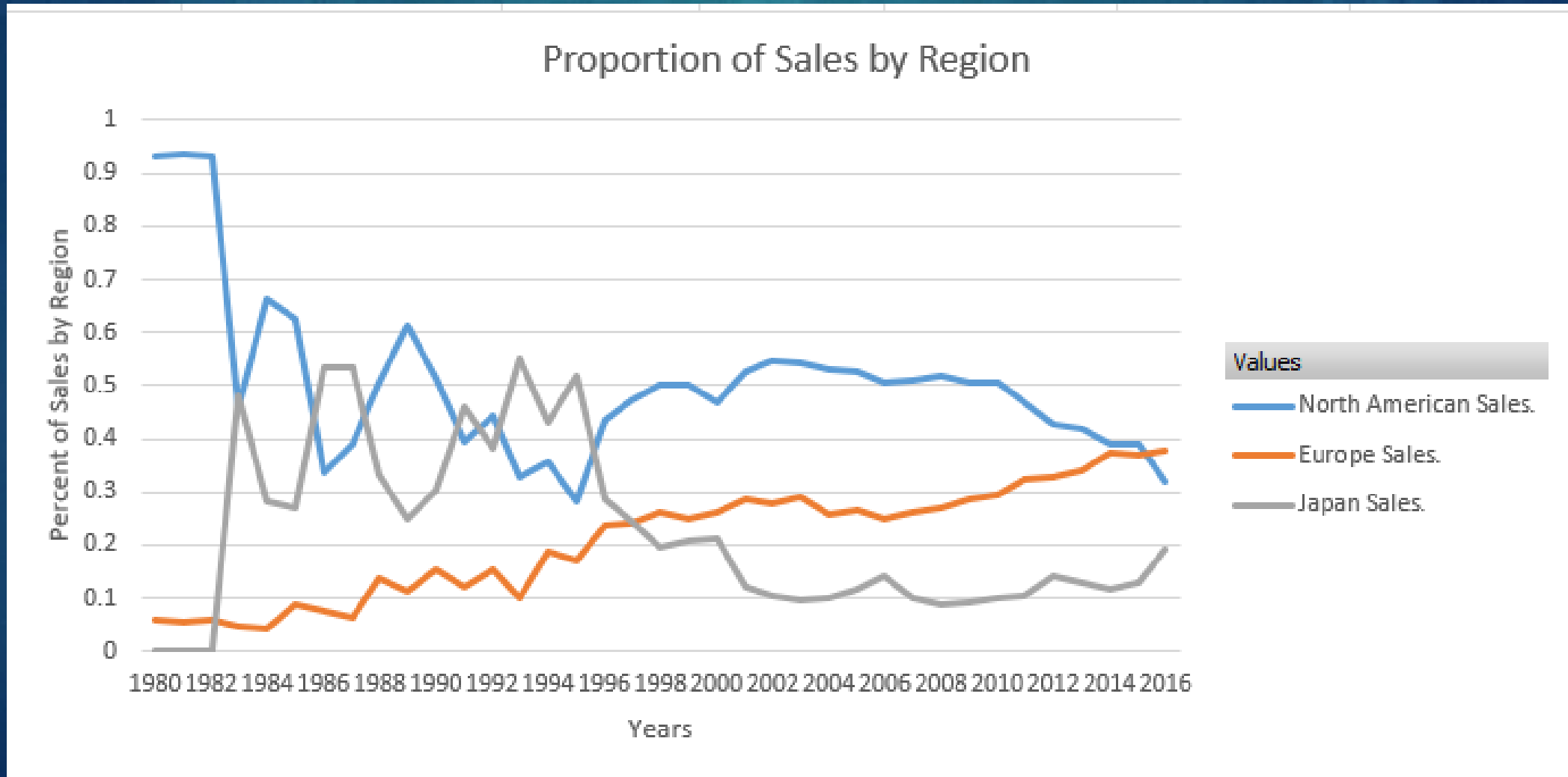
Shooter games is 56% of North America sales and Strategy games are the least with 39% in sales.





# GameCo

The proportion of sales by region between 1980 to 2016 indicates the geographic regions sales expectations vary annually.





# GameCo

7

## Recommendations

- Based on the data GameCo should focus on the top selling games for each region, which are:
  - Best selling games for North America are Shooter, Platform, and Sports.
  - Best selling games for Europe are Racing, Action, and Shooter.
  - Best selling games for Japan are Role Playing, Strategy, and Puzzle.
- Additional research is needed to understand why sales for each of the regions have declined since 2010.
- Redistribute marketing budget to Europe, since they have steady increasing sales and focus on their best-selling games.
- Research into why Europe's sales have been steadily increasing and see if a correlation can be made to help increase sales in North America and Japan.



# Project 2 - Preparing for Influenza Flu Season

## Objective

- Provide information to support staffing plan of medical personnel throughout the United States.
- Prioritize states with vulnerable populations.
- Assess data limitations that may prevent conducting desired analysis.
- Source: CDC, US Census Bureau, Influenza Visits Data Set,

## Key Questions

- Who is affected by the flu?
- Does age make a difference?
- Which state need additional resources?
- Which months are the highest for the flu?

## Tasks

- Cleaning, formatting, address missing data.
- Integrate at least two relevant data sets from different sources.
- Find relationship between variables with at least one correlation.
- Links:  
<https://www.youtube.com/watch?v=ht2oeFXsBWE>
- [https://public.tableau.com/app/profile/rich.bradley/viz/2\\_9Storytellingdashboard2/Story1?publish](https://public.tableau.com/app/profile/rich.bradley/viz/2_9Storytellingdashboard2/Story1?publish)





# Preparing for Influenza Flu Season

The statistical hypothesis being tested are:

- The influenza death rates for people age 65+ is less than or equal to the influenza death rates for people under age 65.
- The influenza deaths for people ages 65+ is greater than the influenza death rates for people under age 65.

t-Test: Two-Sample Assuming Unequal Variances		
	<i>Flu Mortality Ages Under 65</i>	<i>Flu Mortality Ages 65-85+</i>
Mean	0.000269502	0.001310711
Variance	7.63106E-08	2.66235E-07
Observations	459	459
Hypothesized Mean Difference	0	
df	701	
t Stat	-38.11406128	
P(T<=t) one-tail	2.5536E-173	
t Critical one-tail	1.647030228	
P(T<=t) two-tail	5.1072E-173	
t Critical two-tail	1.96335386	

## Influenza Frequency of Deaths by Age Group





# Preparing for Influenza Flu Season

## Next Steps

Preparing for the influenza flu season, the data and research show influenza mortality rates are not going down.

### Recommendations:

The chart on the right is based on the CDC Influenza Deaths data, and I recommend the agency to send resources in **December, January, and February**, which are the highest months for the flu season versus the more mild weather months. I also recommend that the agency send additional resources to the following states:

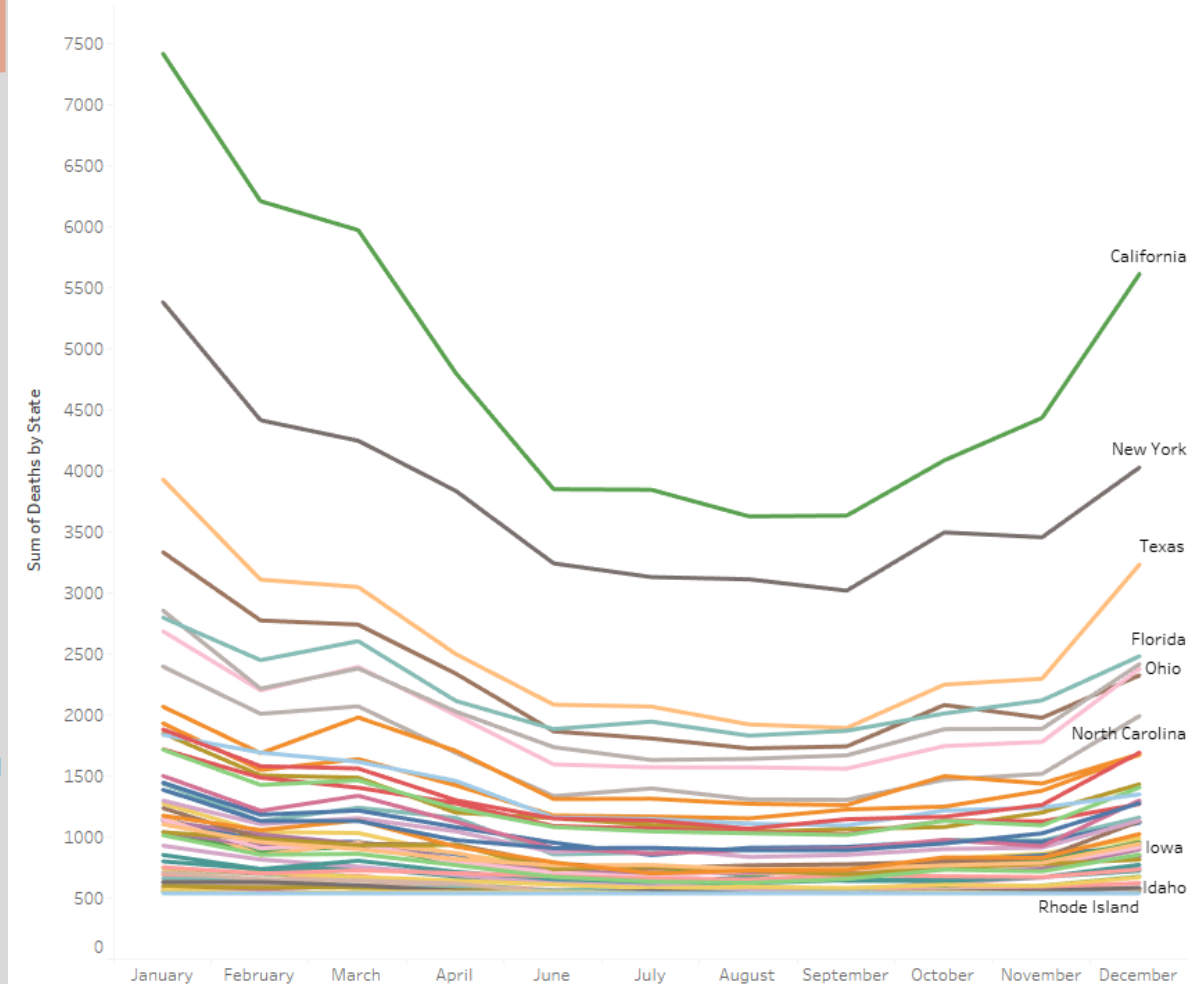
**California, New York, Texas, Florida and Ohio**

**Send resources in November, just prior to the height of the flu season.**

Key Performing Indicators (KPIs) will be used to measure the objectives of this project which is maintaining staff-to-patient ratio of at least 90% and not greater than 110%, when to send staff, and which states to send staff. Monitoring these KPI's will assist the agency's management of their limited resources.

The metrics that will be used are surveys, Twitter, blog posts, and online reviews, prior, during and at the end of the flu season. The responses to key questions will measure performance, patient wellness and duration at the facility and will be analyzed for improvements in flu mortality and help the agency make adjustments regarding there resources and recommendations as needed for future improvements for this project.

Flu Deaths by Month per State





# Project 3 - Rockbuster Data Analysis

## Data Set

- Rockbuster Stealth LLC a movie rental company plans to launch online video rental service to stay competitive.
- Rockbuster's film inventory, customers, payments etc.
- Source: Rockbuster data set

## Key Questions

- What are the key KPI statistics?
- Which countries are Rockbuster customers based in?
- Where are the customers with high value based?

## Tasks

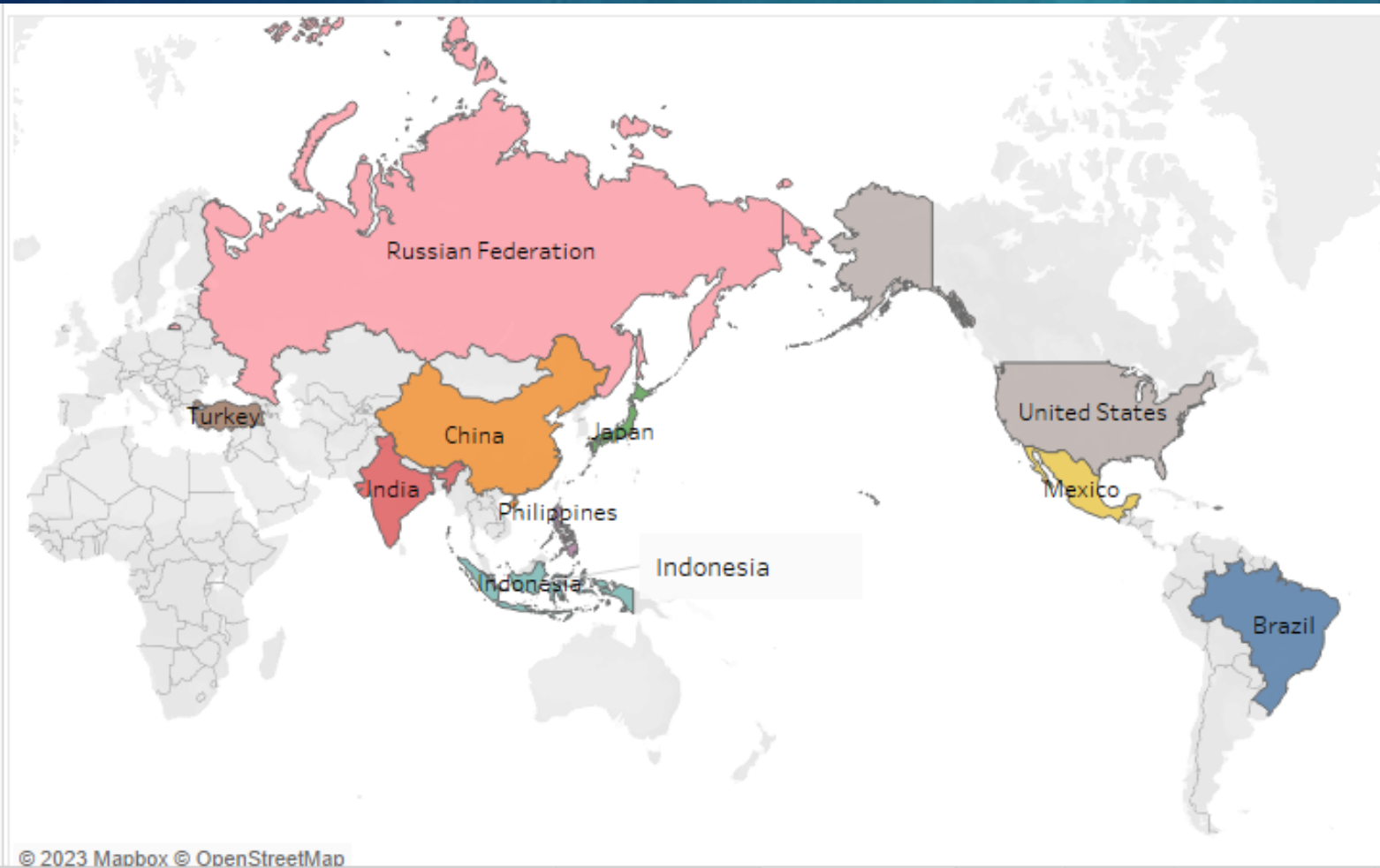
- Create a data dictionary
- Use SQL to create queries, subqueries and CTE's.
- Use Tableau to create visualization.
- Links:
  - [https://public.tableau.com/app/profile/rich.bradley/viz/3\\_10Step1\\_final/top10\\_countries?publish=yes](https://public.tableau.com/app/profile/rich.bradley/viz/3_10Step1_final/top10_countries?publish=yes)
  - [https://public.tableau.com/app/profile/rich.bradley/viz/5\\_8RockBustertop5customers/Top5Customers?publish=yes](https://public.tableau.com/app/profile/rich.bradley/viz/5_8RockBustertop5customers/Top5Customers?publish=yes)
  - <https://github.com/Rich-Brad/sql-rockbuster-movie-rental>





# RockBuster Data Analysis

## The Top Ten Countries Brand Focus

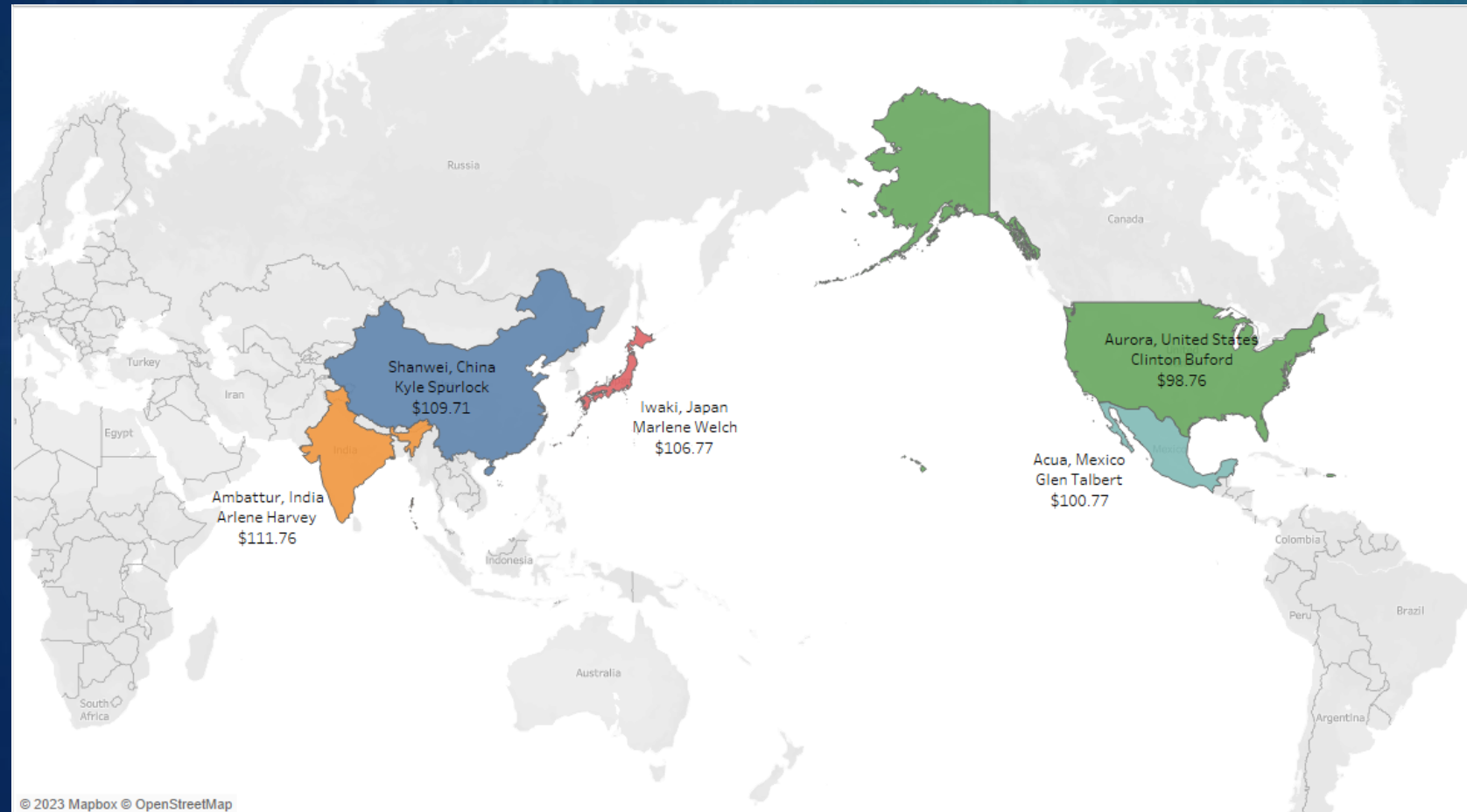


```
1 SELECT country,  
2     COUNT(*)  
3 FROM customer A  
4 INNER JOIN address B ON A. address_id = B. address_id  
5 INNER JOIN city C ON B. city_id = C. city_id  
6 INNER JOIN country D ON C. country_id = D. country_id  
7 GROUP BY country  
8 ORDER BY COUNT(*) DESC  
9 LIMIT 10;  
10 |
```



# RockBuster Data Analysis

## The Top Five Customers Based on Sales



```
1 SELECT A.customer_id AS cust_id,SUM(amount) AS amount,
2       B.first_name,
3       B.last_name,
4       D.city,
5       E. country
6 FROM payment A
7 INNER JOIN customer B ON A.customer_id = B. customer_id
8 INNER JOIN address C ON B. address_id = C. address_id
9 INNER JOIN city D ON C. city_id = D. city_id
10 INNER JOIN country E ON D. country_id = E. country_id
11 WHERE city IN ('Aurora', 'Acua', 'Citrus Heights', 'Iwaki', 'Ambattur',
12              'Shanwei', 'So Leopoldo', 'Teboksary', 'Tianjin', 'Cianjur')
13 GROUP BY cust_id, first_name, last_name,city, country
14 ORDER BY SUM(amount) DESC
15 LIMIT 5;
```



# RockBuster Data Analysis

## RockBuster Recommendations

- Rockbuster should leverage the top five customers in their marketing and image branding efforts to increase market share in the top ten countries and top cities.
- In addition, Rockbuster, through their marketing and image branding can show they are a global company. This could help market expansion for future projects.
- Finally, Rockbuster will need to continue testing and analyzing research results to better understand how to improve and build on their marketing and image branding in the top ten countries and cities.





# Project 4 - Instacart Grocery Basket Analysis

## Objective

- Online grocery store that operates through an app.
- Want to uncover information about their sales patterns.
- They want to target different customers with applicable marketing campaigns.
- Source: "The Instacart Online Grocery Shopping Dataset 2017", Accessed from [www.instacart.com/datasets/grocery-shopping-2017](https://www.instacart.com/datasets/grocery-shopping-2017) via [Kaggle](https://www.kaggle.com) on 7/1/2023.

## Key Questions

- What is the busiest hours of the day?
- What is the busiest day of the week?
- What is the frequency of customers returning to Instacart?
- What is the age and income of the customers?

## Tasks

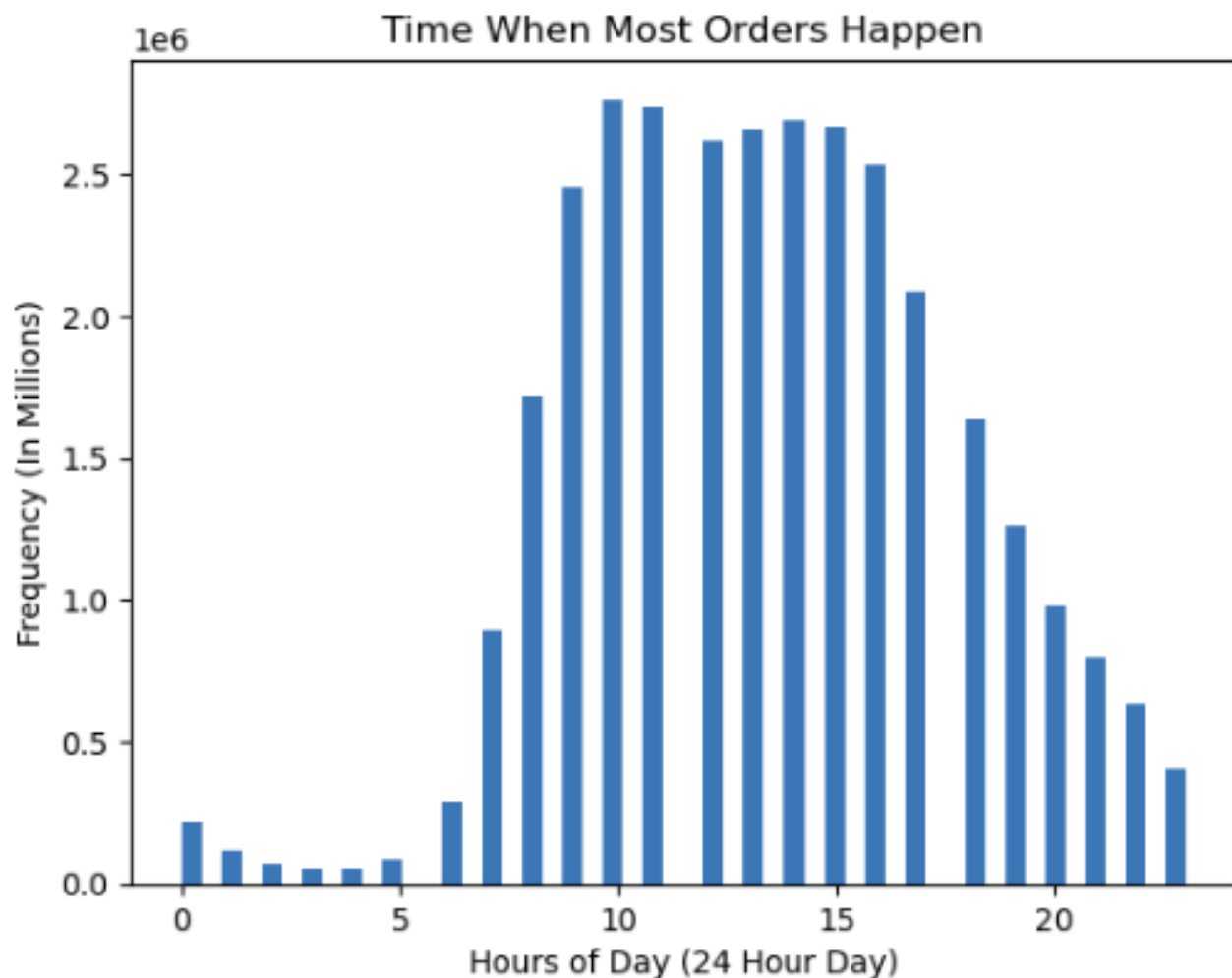
- Consistency check for missing values, duplication, and fix any mixed-type variables.
- Used Python libraries and Jupyter to wrangle grouping and aggregating data.
- Created visualizations with Python.
- Links:
  - <https://github.com/Rich-Brad/python-instacart-analysis>





# Instacart Grocery Basket Analysis

This histogram shows the bulk of orders happen between 8am to 4pm.

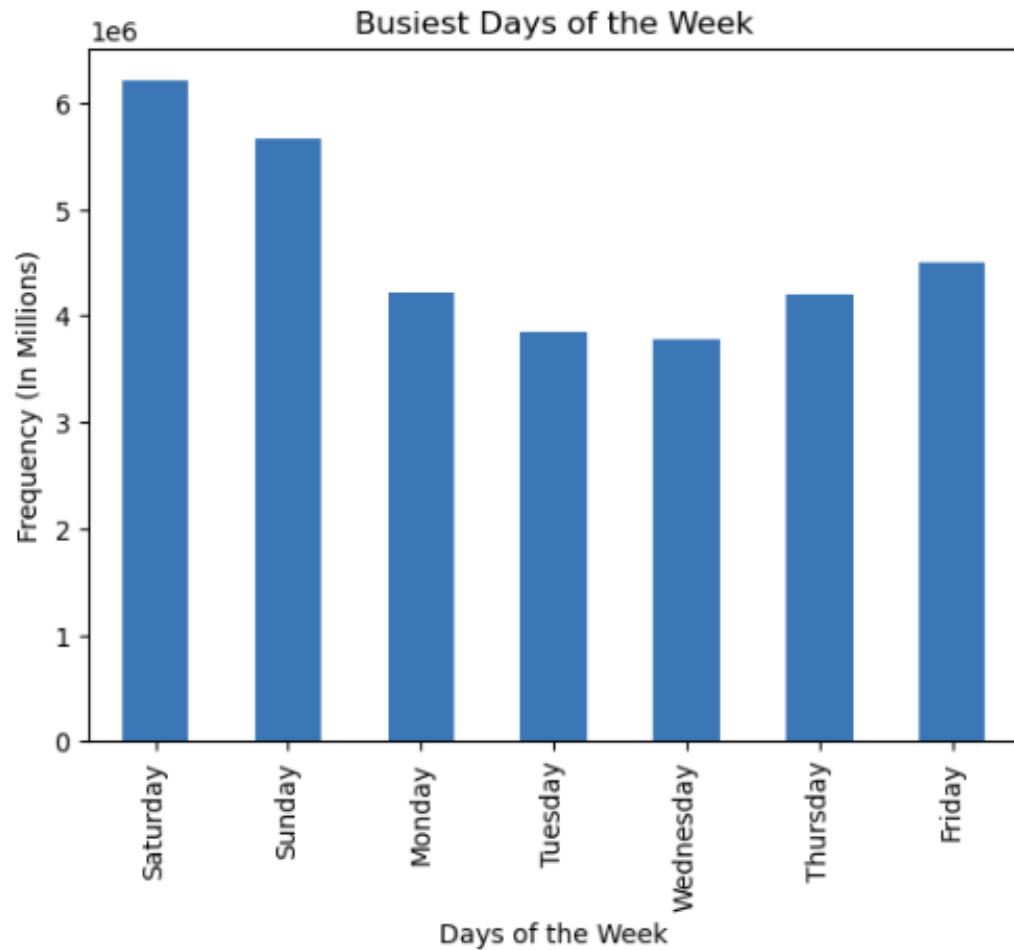


```
hist_2 = cust_ords_prods_merge['order_hour_of_day'].plot.hist(bins = 50)
plt.xlabel("Hours of Day (24 Hour Day)")
plt.ylabel("Frequency (In Millions)")
plt.title("Time When Most Orders Happen")
```



# Instacart Grocery Basket Analysis

The busiest day of the week is Saturday with Tuesday and Wednesday being the least busy days.

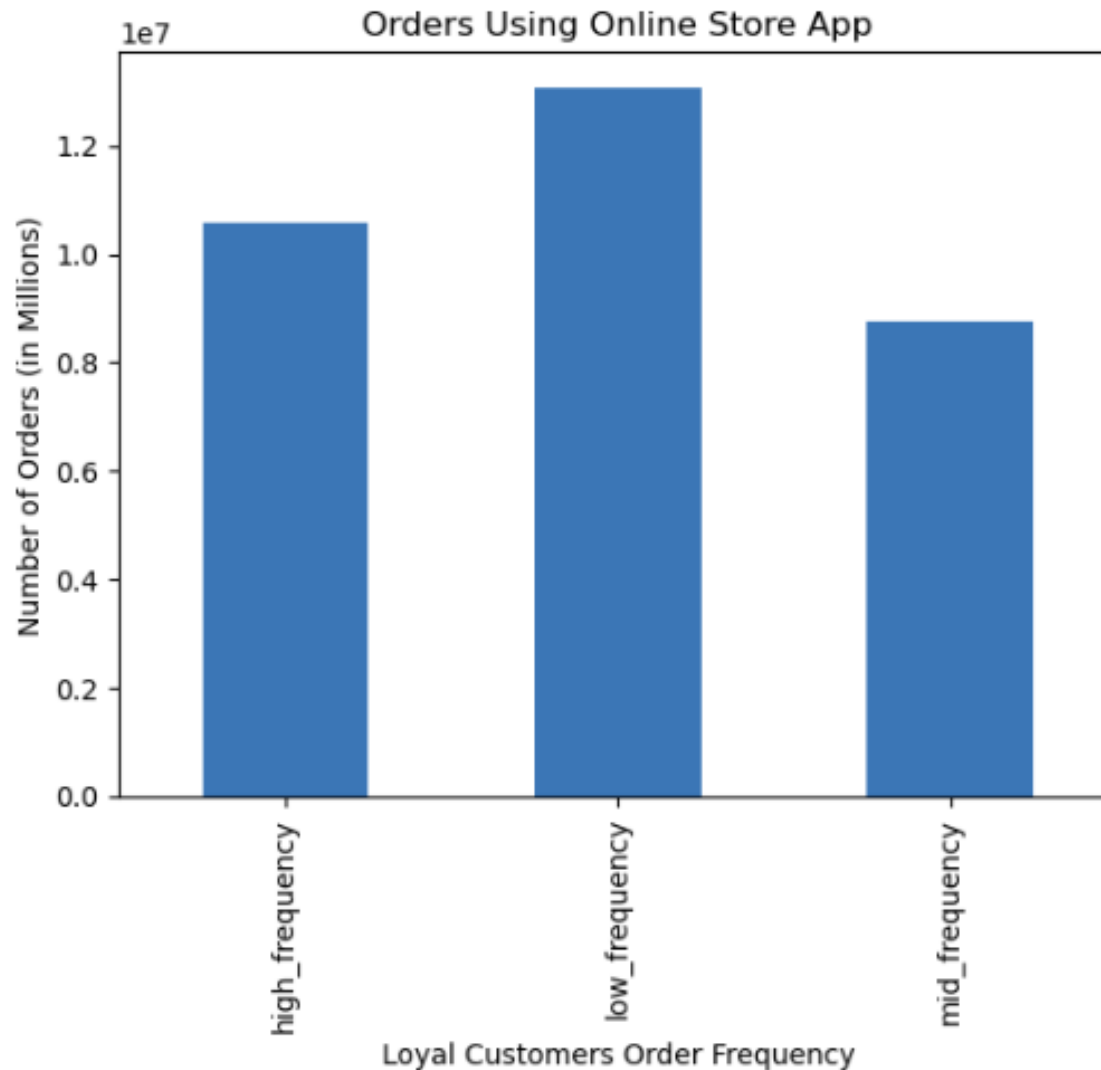


```
bar = cust_ords_prods_merge['orders_day_of_the_week'].value_counts().sort_index().plot.bar()
plt.xticks([0,1,2,3,4,5,6],["Saturday", "Sunday", "Monday", "Tuesday", "Wednesday", "Thursday", "Friday"])
plt.xlabel('Days of the Week')
plt.ylabel('Frequency (In Millions)')
plt.title('Busiest Days of the Week')
```



# Instacart Grocery Basket Analysis

Customer loyalty by frequent use of the app.



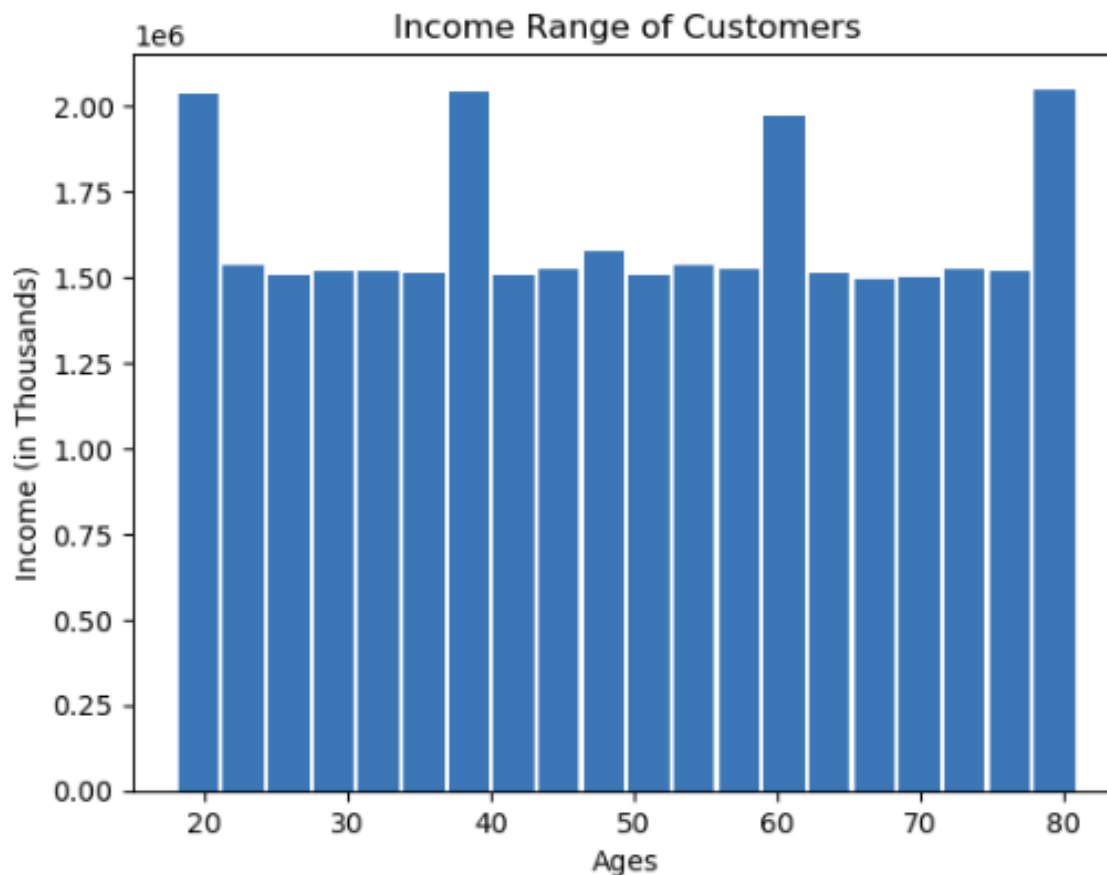
```
bar_2 = cust_ords_prods_merge['loyalty_flag'].value_counts().sort_index().plot.bar()
plt.xlabel('Loyal Customers Order Frequency')
plt.ylabel('Number of Orders (in Millions)')
plt.title('Orders Using Online Store App')
```



# Instacart Grocery Basket Analysis

This histogram shows most incomes are about 200,000 and below in the different age groups.

With the highest concentration of incomes at 150,000.



```
cust_ordr_prods_merge.hist(column = "Age", bins= 20, rwidth = .9, grid = False)
plt.xlabel('Ages')
plt.ylabel('Income (in Thousands)')
plt.title('Income Range of Customers')
```





# Instacart Grocery Basket Analysis

## Recommendations

- The least busy days are Tuesday and Wednesday. Knowing this the marketing team can provide incentives or promotions to increase sales on those slower days.
- The fewest orders happen between 11pm and 6am. The marketing team could do a midnight special or early bird shopping special to encourage more sales during these hours.
- Have the marketing and sales team use incentives, for example, coupons or discounts to help increase customer use of the online app more and decrease the non-frequent customers to becoming regular or frequent customers.
- Find ways to reward the loyal customers and incentivize the new customers with specials based on their purchases.

# Project 5 - Flight Delay Project Analysis

## Objective

- Determine the flight delays **greater** than 15 minutes
- **Which** airlines are impacted the most
- Provide information and insight regarding flight delays throughout the United States.
- Want to uncover which airports and airlines have the greatest delays.
- Access which delays, if any, have a correlation.
- Source: Open Source: Kaggle.com  
Original Author Eugeny Osetrov (Owner)
- Time Series Analysis Data.gov  
Publisher opendata.Maryland.gov  
Maintainer Christine Howard

## Key Questions

- Is there a connection between certain airlines and airports that show a higher delay rate within the data set?
- What airline has the worse delays, greater than 15 minutes?
- Are there certain airlines that are consistently having higher delays than other airlines?
- Are there certain airports that are consistently having higher delays than other airports?

## Tasks

- Consistency check for missing values, duplication, and fix any mixed-type variables.
- Used Python libraries and Jupyter to wrangle grouping and aggregating data.
- Created visualizations with Python.
- Create a data dashboard with Tableau.
- Links:
  - <https://github.com/Rich-Brad/Airline-Delays.git>

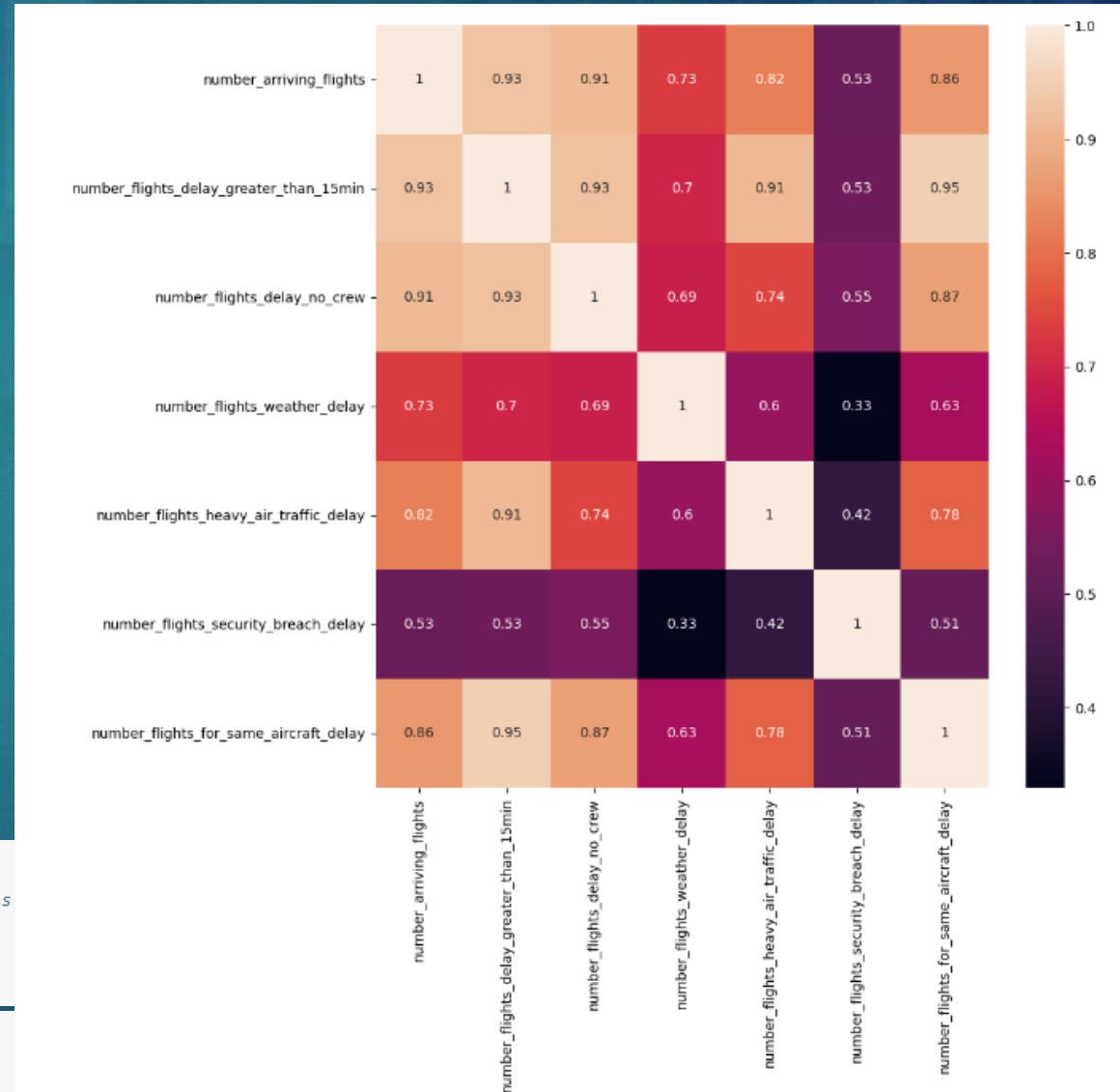
# Flight Delay Correlation Matrix

This heatmap shows the correlation of the different delay categories and how strong or weak the correlations are in relationship to each delay.

- The coefficient range between .3 - .5, the darker purple colors show a weak relationship.
- The coefficient range between .5 – .7 the purple to red colors shows a moderate relationship.
- The coefficient range between .7 – 1.0, the darker beige to light beige colors shows a strong relationship between the variables.

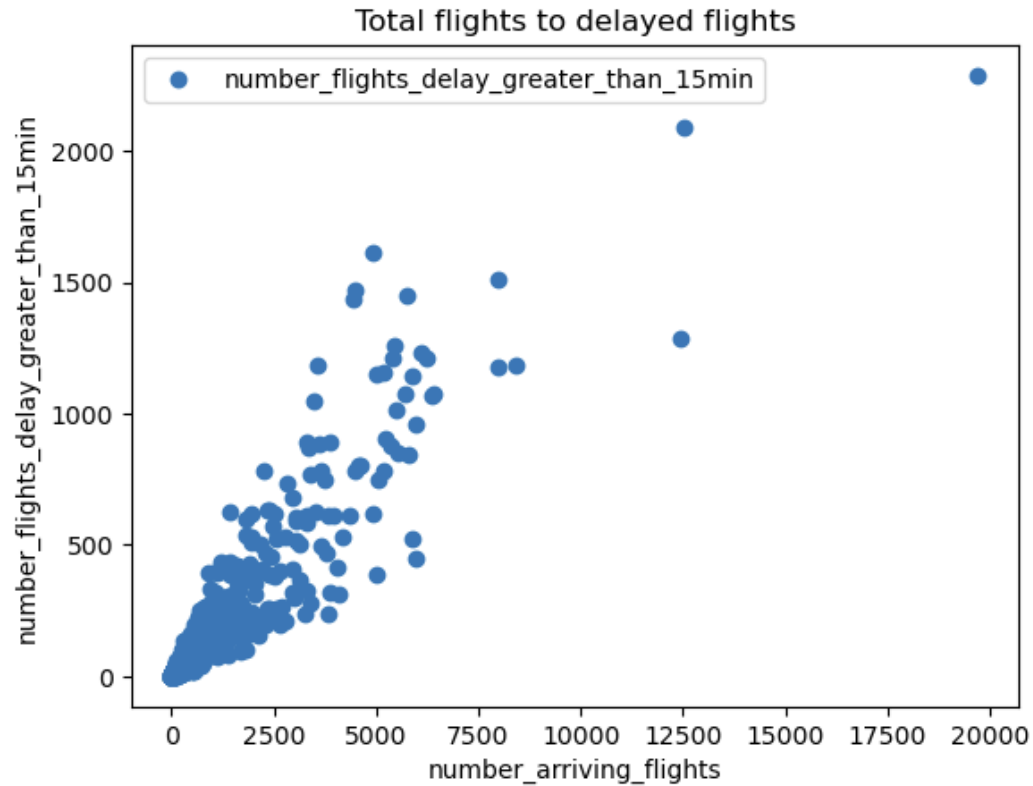
```
f = plt.figure(figsize=(15, 15)) # figure size
plt.matshow(df_air_delay_1.corr(), fignum=f.number) # type of plot
plt.xticks(range(df_air_delay_1.shape[1]), df_air_delay_1.columns, fontsize=14, rotation=45) # x axis labels
plt.yticks(range(df_air_delay_1.shape[1]), df_air_delay_1.columns, fontsize=14) # y axis labels
cb = plt.colorbar() # add a colour legend (called colorbar)
cb.ax.tick_params(labelsize=14) # add font size
plt.title('Airline Delay Correlation Matrix', fontsize=14) # add title
```

```
plt.figure(figsize = (10,10))
sns.heatmap(df_air_delay_1.corr(), annot = True)
```



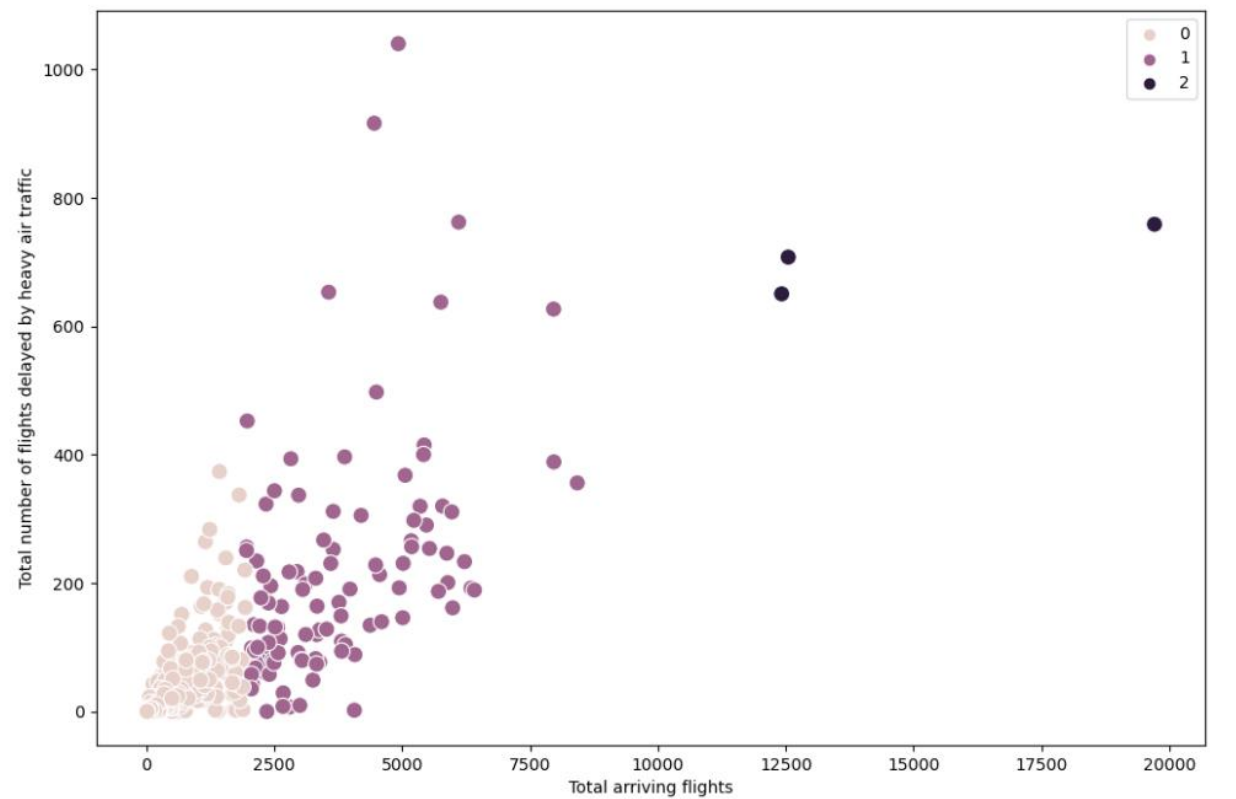


# Flight Delay Normalization and Scatterplots



```
# Create a scatterplot
```

```
sns.scatterplot(x = 'number_arriving_flights', y = 'number_flights_delay_greater_than_15min', data = df_air_dlay_1)
```



```
plt.figure(figsize=(12,8))
ax = sns.scatterplot(x=df_air_dlay['number_arriving_flights'], y=df_air_dlay['number_flights_heavy_air_traffic_delay'], hue=kmean)
ax.grid(False) # This removes the grid from the background.
plt.xlabel('Total arriving flights') # Label x-axis.
plt.ylabel('Total number of flights delayed by heavy air traffic') # Label y-axis.
plt.show()
```



# Flight Delays By Airport

25

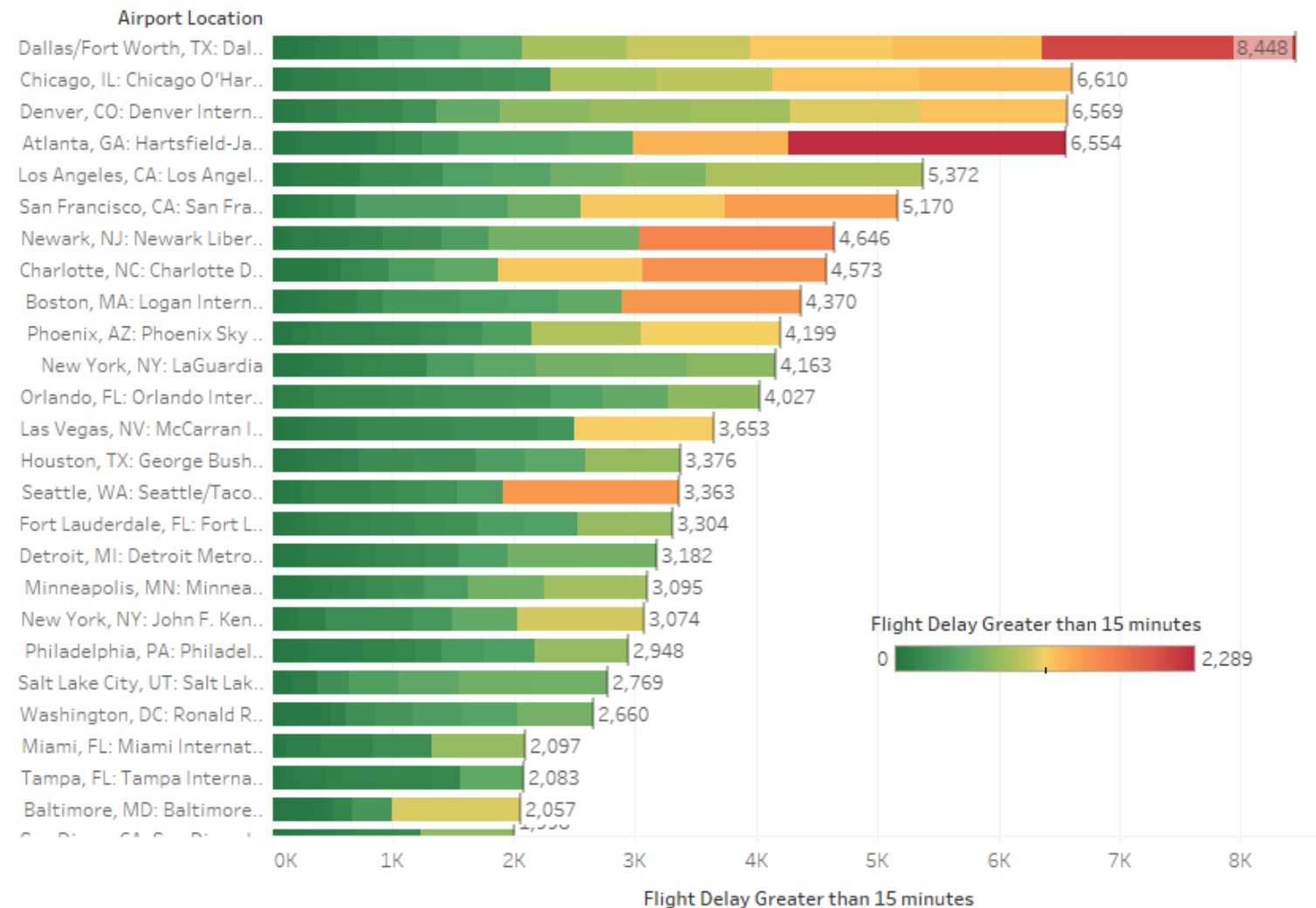
## Hypothesis

“If arriving flights encounter delays, then the number of arriving flight delays will be 15 minutes or longer”

## Project Question

“Of the flight delays greater than 15 minutes, flights for the same aircraft, heavy air traffic, and weather delays, which has the most impact”?

Delays Greater than 15 Minutes by Airport

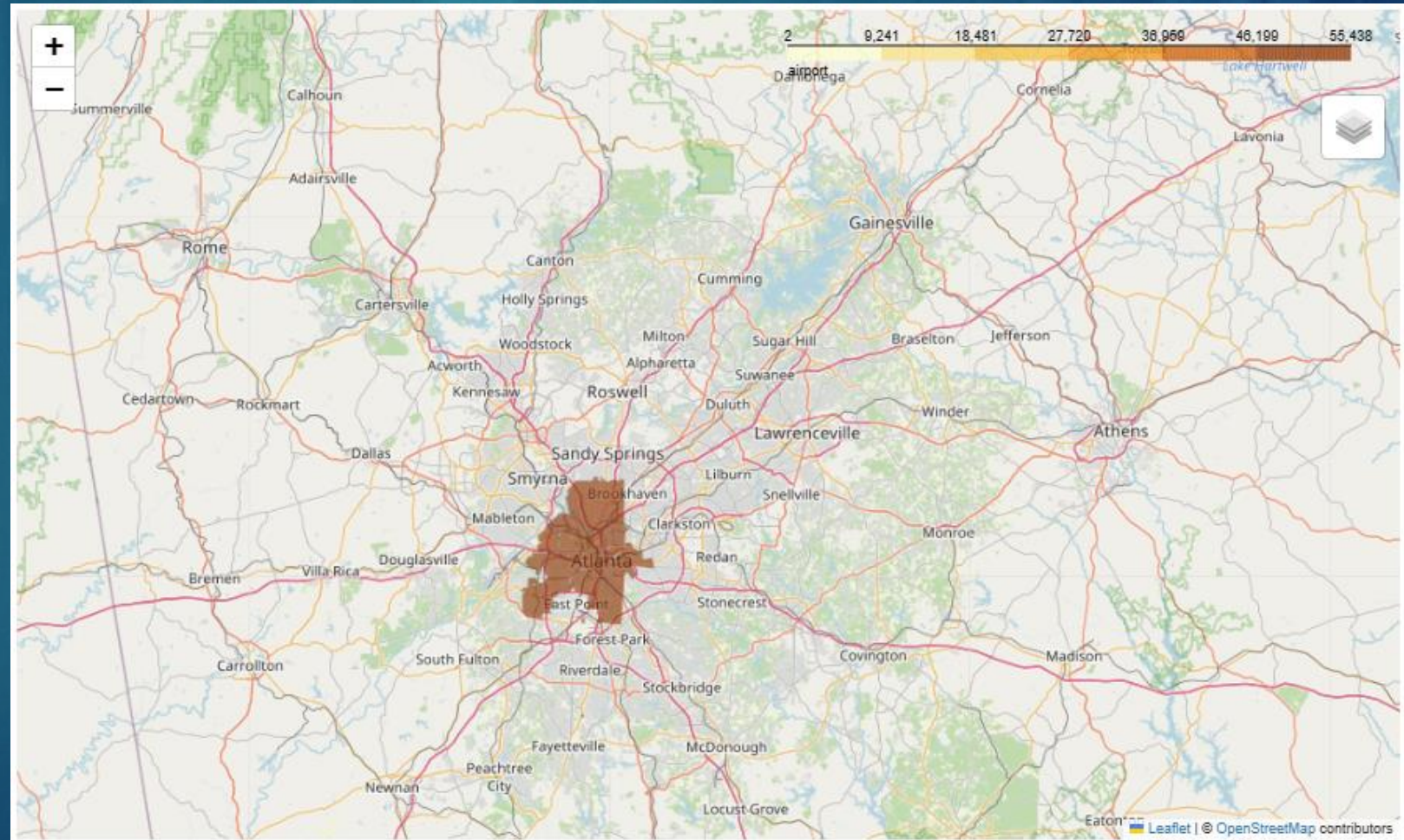


# Flight Delay Choropleth Map of US Airports

26

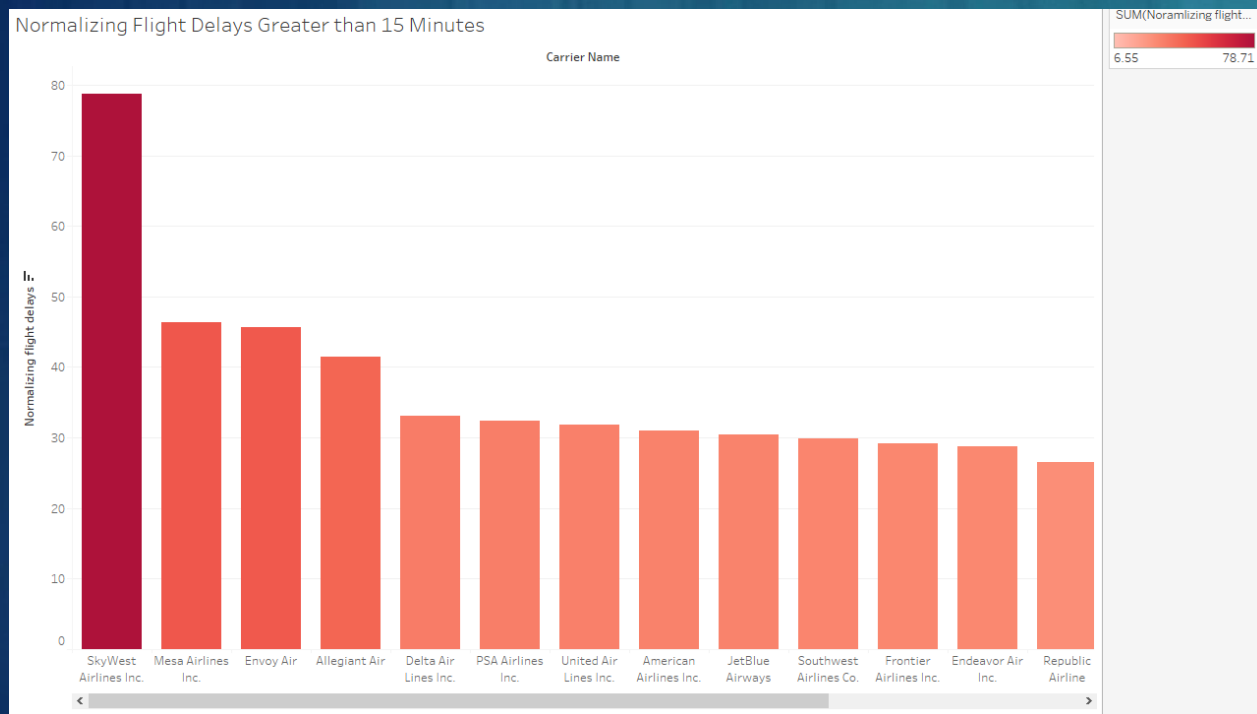
This analysis shows the cities where the airports are located.

In this analysis Atlanta has the most with 55,438 arriving flights.



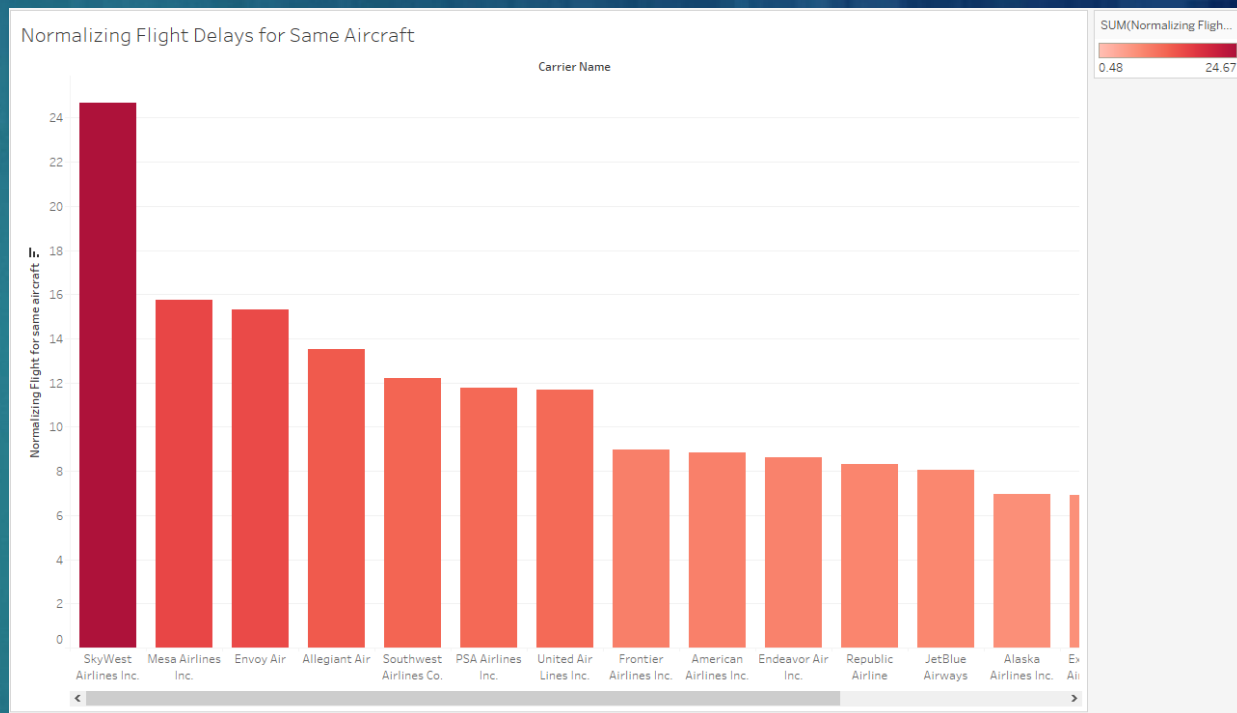
# Flight Delay Airlines Affected By Delays

Normalizing Flight Delays Greater than 15 Minutes



SkyWest Airlines Inc. is affected most with delays 15 minutes and greater. Followed by Mesa Airlines Inc. and Envoy Air.

Normalizing Flight Delays for Same Aircraft



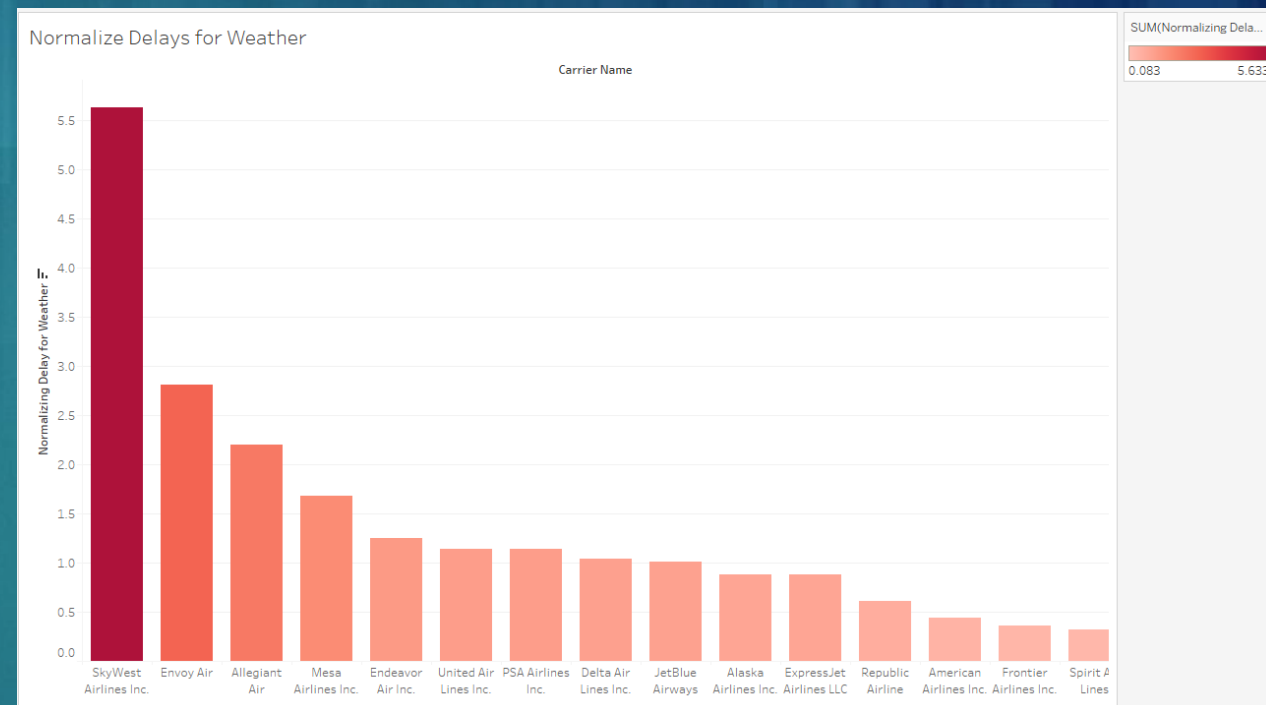
SkyWest Airlines Inc. is most affected by flight delays for same aircraft. Followed by Mesa Airlines Inc. and Envoy Air.



# Flight Delay Airlines Affected By Delays



Envoy Air is most affected by heavy air traffic. Followed by Allegiant air and Republic Airline.



SkyWest Airlines Inc is most affected by weather delays. Followed by Envoy Air and Allegiant Air.



# Recommendations and Next steps

29

Preparing for flights in December, the data shows there is a strong correlation if a flight is going to be delayed, the delay will be longer than 15 minutes.

The hypothesis, "If arriving flights encounter delays, then the number of arriving flight delays will be 15 minutes or longer". (See chart on the right).

## Recommendations:

Of the flight delays greater than 15 minutes the air lines most affected:

\* Flights for the same aircraft delay

Sky West, Envoy, and Mesa Airlines

\* Heavy air traffic and weather delays

Envoy, Allegiant, and Delta Airlines

\* Weather delays

Sky West, Envoy, and Allegiant Airlines

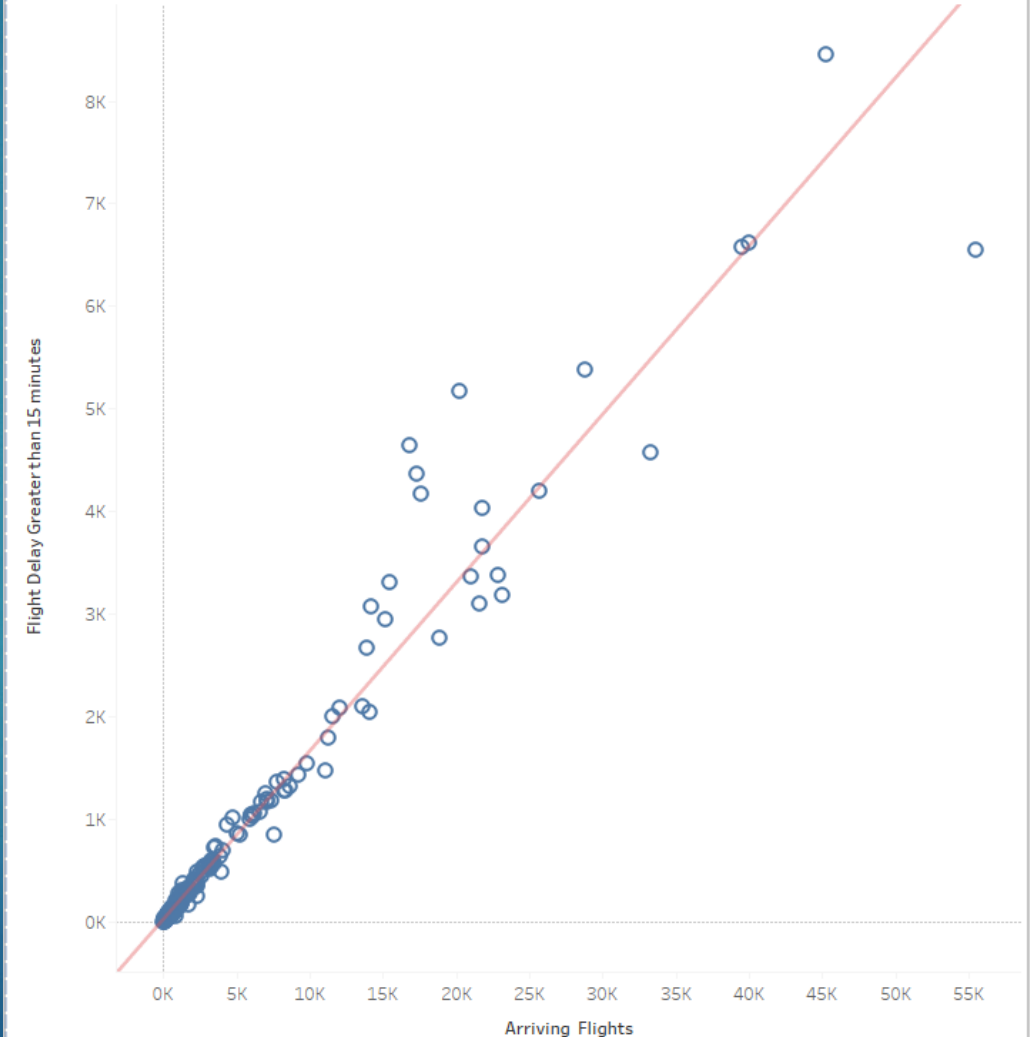
The management of all the airlines should be proactive in researching the reason these delays are occurring. Although weather delays may be a challenge, consideration should be taken to better understand where those flights and airports are to better plan and provide travelers a better experience.

Also, airline companies should utilize surveys, twitter, blog posts, online reviews etc. to gain insight from travelers. The analysis from those responses should be available to the stakeholders as needed to assess and make informed decisions to improve the overall experience of their customers.

Finally, we need to continue to analyze and assess future December flight delay data to measure any improvement in flight delays, and adjust or make additional recommendations as needed.

Author - Rich Bradley

Trend Line Arriving Flights and Flight Delays 15 Minutes and Greater



## Sources:

Open Source: Kaggle.com  
Original Author Eugeniy Osetrov (Owner)

Time Series Analysis  
Data.gov Publisher opendata.Maryland.gov  
Maintainer Christine Howard

## Links:

<https://www.youtube.com/watch?v=ht2oeFXsBWE>

[https://public.tableau.com/app/profile/rich.bradley/viz/2\\_9Storytellingdashboard2/Story1?publish](https://public.tableau.com/app/profile/rich.bradley/viz/2_9Storytellingdashboard2/Story1?publish)

[https://public.tableau.com/app/profile/rich.bradley/viz/3\\_10Step1\\_final/top10countries?publish=yes](https://public.tableau.com/app/profile/rich.bradley/viz/3_10Step1_final/top10countries?publish=yes)

[https://public.tableau.com/app/profile/rich.bradley/viz/5\\_8RockBustertop5customers/Top5Customers?publish=yes](https://public.tableau.com/app/profile/rich.bradley/viz/5_8RockBustertop5customers/Top5Customers?publish=yes)

<https://github.com/Rich-Brad/sql-rockbuster-movie-rental>

<https://github.com/Rich-Brad/python-instacart-analysis>

<https://github.com/Rich-Brad/Airline-Delays.git>

[https://public.tableau.com/app/profile/rich.bradley/viz/Ex6\\_7A/AirlineFlightDelays?publish=yes](https://public.tableau.com/app/profile/rich.bradley/viz/Ex6_7A/AirlineFlightDelays?publish=yes)

<https://www.kaggle.com/datasets/eugeniyosetrov/airline-delays/data>

<https://catalog.data.gov/dataset/mva-vehicle-sales-counts-by-month-for-calendar-year-2002-2020-up-to-october>

Rich Bradley

Conclusion Slide

Data Analytics Case Studies

Contact me with any questions, or if you would like to collaborate with me at [rbbrad@gmail.com](mailto:rbbrad@gmail.com)