RUSAMIJAN PERMISON

Data Analytics Portfolio

Email: meeps.analyst@gmail.com
Portfolio: www.meepermison.com

PROJECTS

Analyze sales data and other variables using Python , Excel and Tableau Public (Github)

Analyze sales and customer profiling using Python, Excel, Tableau Public and (Github)

Identify insights and customer analysis by using SQL , Tableau Public and (Github)

Identify flu season and staffing needs in the US by using Excel and Tableau Public

Analyze global videogame retail sales by using Excel, PowerPoint and Tableau Public

US CAFE’ STORES SALES DATA ANALYSIS

OBJECTIVE

Discover relationships between Sales and other variables. Which variable is the most effective to Sales

DATA

This data is publicly available open-source data. It was downloaded from Kaggle.com (US CAFÉ’ Stores Sales)

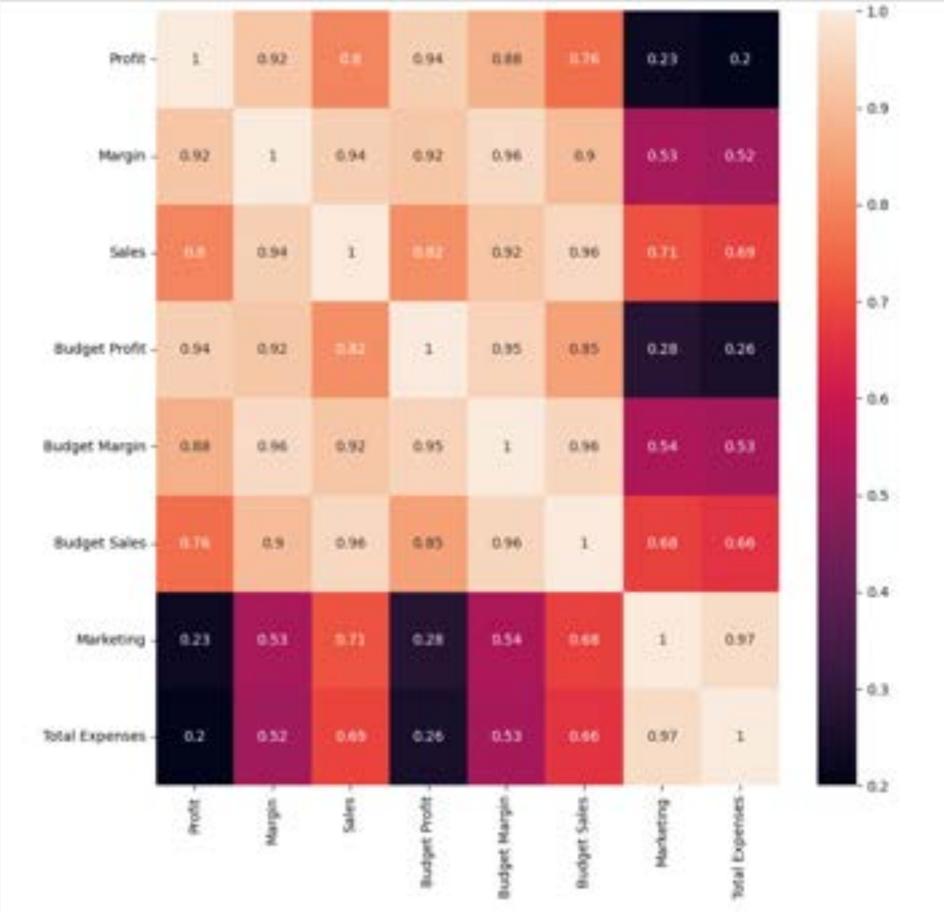
TOOLS & SKILLS

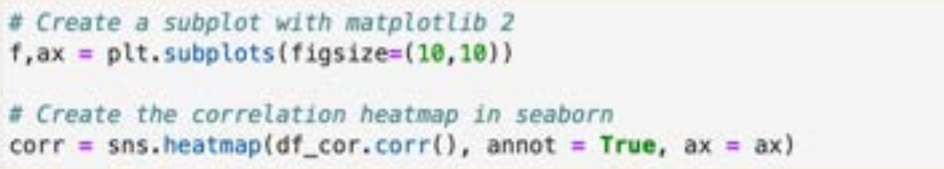
- Data cleaning (wrangle, consistency checks)
- Data manipulation (grouping, aggregating, subsetting, exporting)
- Advanced analysis(geospatial analysis, linear regression analysis, clustering analysis)

LIMITATION

- The data contains only 2020 to 2021
- Dataset contains only 20 States In the US

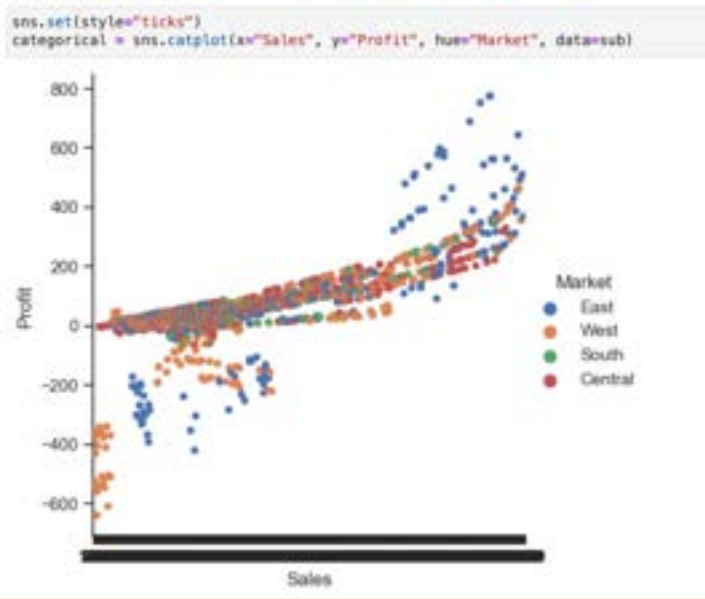
EXPLORING RELATIONSHIPS





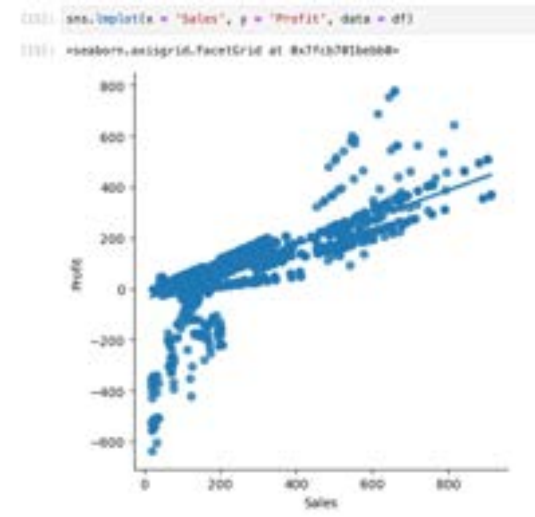
```
# Create a subplot with matplotlib 2
f,ax = plt.subplots(figsize=(10,10))

# Create the correlation heatmap in seaborn
corr = sns.heatmap(df_cor.corr(), annot = True, ax = ax)
```

A correlation heatmap helps verify through colors the strength of relationships between all variables. The relationship between Sales and Profit is 0.8 – it is medium positive relationship.



```
sns.set(style="ticks")
categorical = sns.catplot(x="Sales", y="Profit", hue="Market", data=sub)
```

A Categorical Plot shows East and West are the highest Sales?



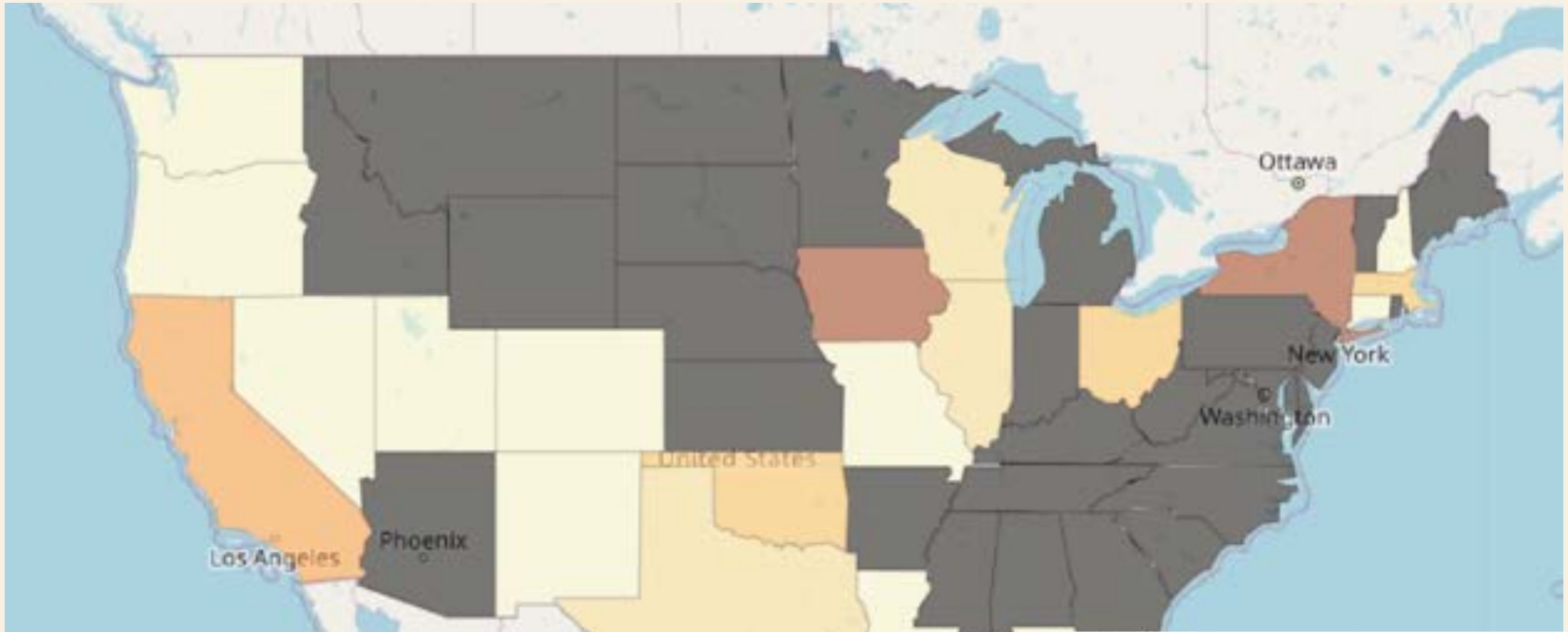
```
[11]: sns.lmplot(x = 'Sales', y = 'Profit', data = df)
[11]: <seaborn.axisgrid.FacetGrid at 0x7fcb701beb0b0>
```

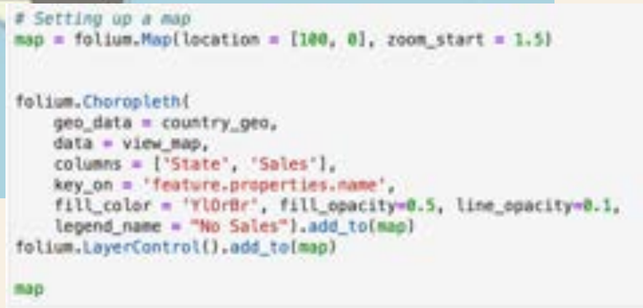
Scatterplot shows the relationship between Sales vs Profit with the line plot. There is positive relationship.

Hypothesis:

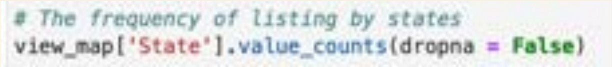
Which factors have the most affect on sales in the US?

GEOSPATIAL ANALYSIS

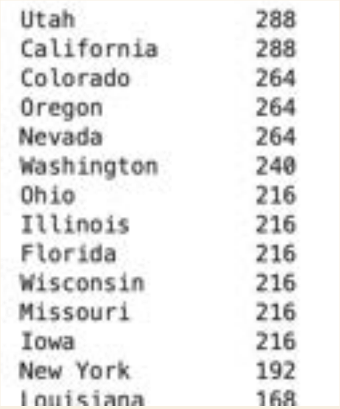




```
# Setting up a map
map = folium.Map(location = [100, 0], zoom_start = 1.5)


folium.Choropleth(
    geo_data = country_geo,
    data = view_map,
    columns = ['State', 'Sales'],
    key_on = 'feature.properties.name',
    fill_color = 'YlOrBr', fill_opacity=0.5, line_opacity=0.1,
    legend_name = "No Sales").add_to(map)
folium.LayerControl().add_to(map)

map
```



```
# The frequency of listing by states
view_map['State'].value_counts(dropna = False)
```

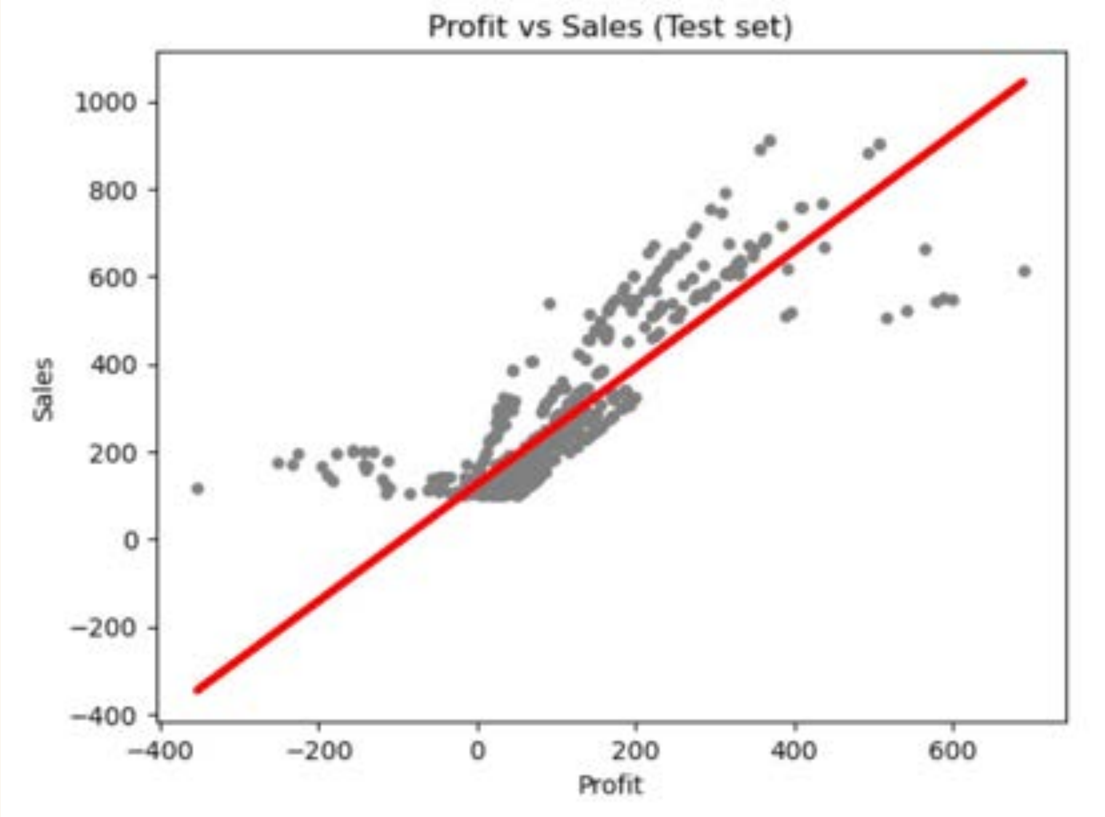


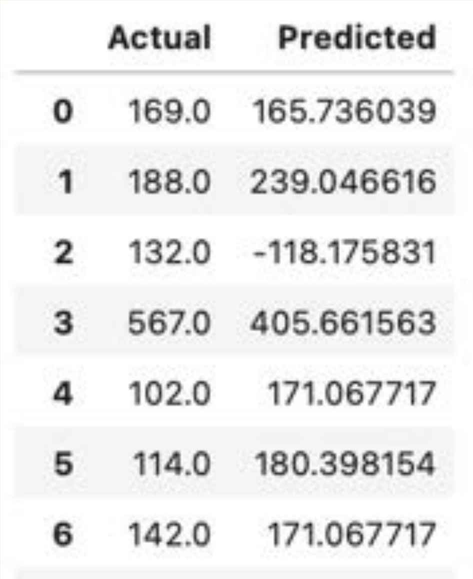
```
Utah            288
California      288
Colorado        264
Oregon          264
Nevada          264
Washington      240
Ohio            216
Illinois        216
Florida         216
Wisconsin       216
Missouri        216
Iowa            216
New York        192
Louisiana       168
```

Questions to explore:
Which State has the most Sales (From 20 States)? From California(288), Colorado(264), Washington(240) and Ohio(214)

Do States have an impact on the Sales amount? yes, States has an impact on the sales amount.

ADVANCED TECHNIQUES-REGRESSION

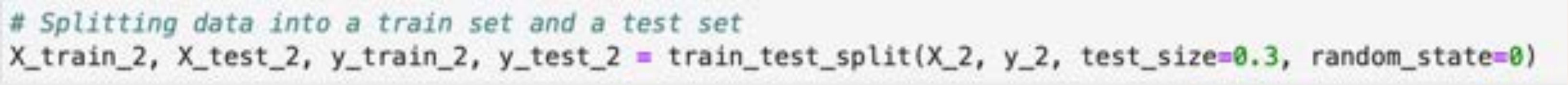




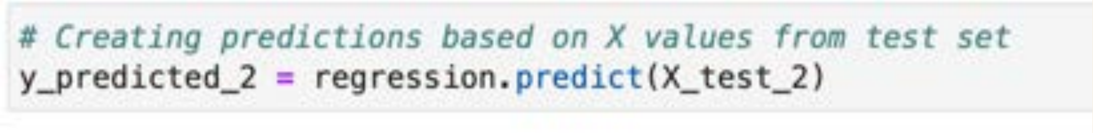
	Actual	Predicted
0	169.0	165.736039
1	188.0	239.046616
2	132.0	-118.175831
3	567.0	405.661563
4	102.0	171.067717
5	114.0	180.398154
6	142.0	171.067717

Slope: [[1.33291958]]
Mean squared error: 8107.694544475579
R2 score: 0.649888432083068

There is a strong positive relationship between variables. The number that represents the sales increase then the profit also increase. The high MSE and high R2 score are good for making predictions

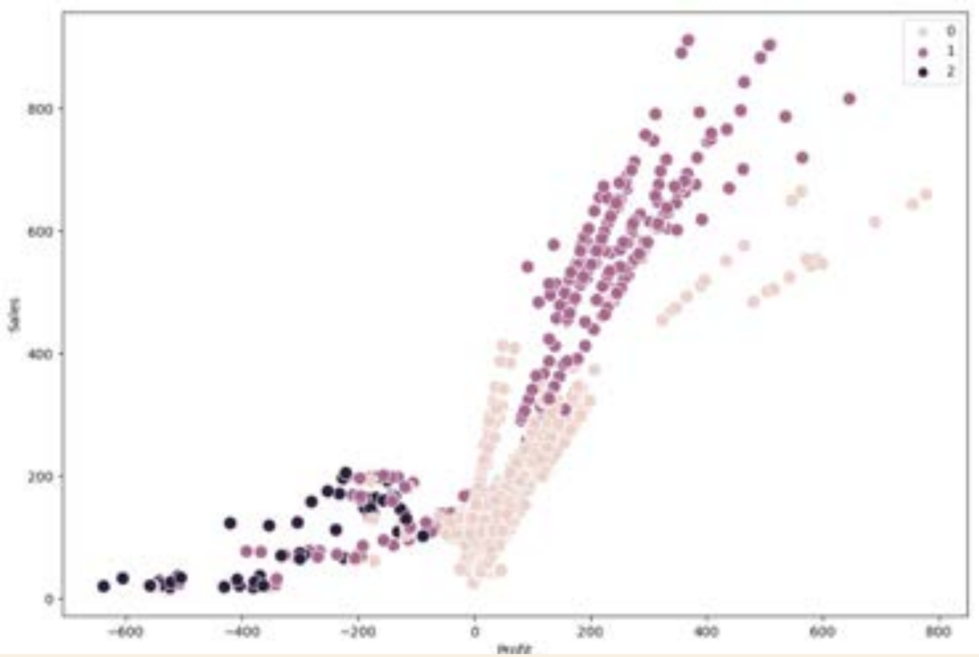


```
# Splitting data into a train set and a test set
X_train_2, X_test_2, y_train_2, y_test_2 = train_test_split(X_2, y_2, test_size=0.3, random_state=0)
```

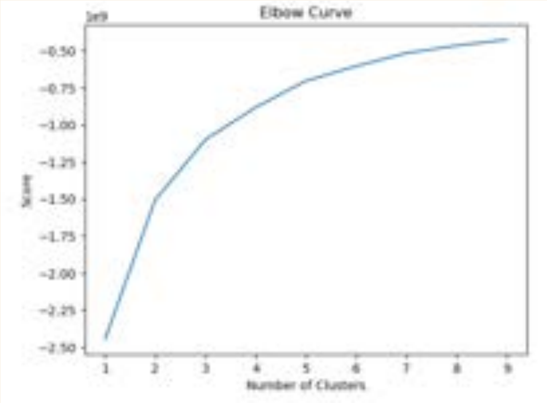


```
# Creating predictions based on X values from test set
y_predicted_2 = regression.predict(X_test_2)
```

ADVANCED TECHNIQUES-CLUSTERING

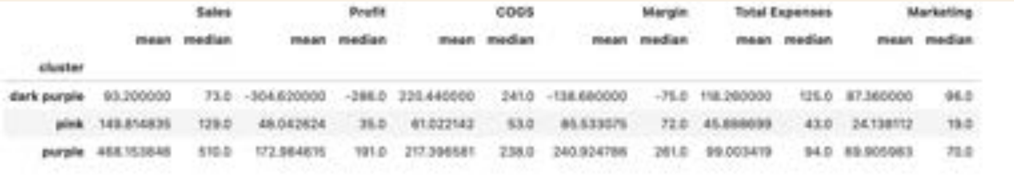


The first cluster, in medium purple (coded as “1” in the legend), contains the points with the highest profit and the highest sales. The second cluster, in dark purple (coded as “2” in the legend), It gathers the data points with lowest profit and relatively lowest sales. The third cluster, in pink (coded as “0” in the legend), includes points with high profit and high sales but less than the first cluster.



Elbow technique. The optimal number of clusters shouldn’t be too many (otherwise, there won’t be much difference between them), while also not being too few. What the elbow technique does, then, is show you the breaking point which adding more clusters won’t help better explain the variances in your data.

Descriptive analysis for clustering



	Sales		Profit		COGS		Margin		Total Expenses		Marketing	
	mean	median	mean	median	mean	median	mean	median	mean	median	mean	median
cluster												
dark purple	93.200000	73.0	-304.620000	-286.0	220.440000	241.0	-138.680000	-75.0	118.260000	125.0	87.360000	96.0
pink	149.814835	129.0	48.042624	35.0	61.022142	53.0	85.533075	72.0	45.888099	43.0	24.138112	19.0
purple	488.153846	510.0	172.984615	191.0	217.396581	238.0	240.924786	261.0	99.003419	94.0	69.905983	73.0

•The purple cluster has the best stats in almost all categories. Sales and Profit are highest in mean and median

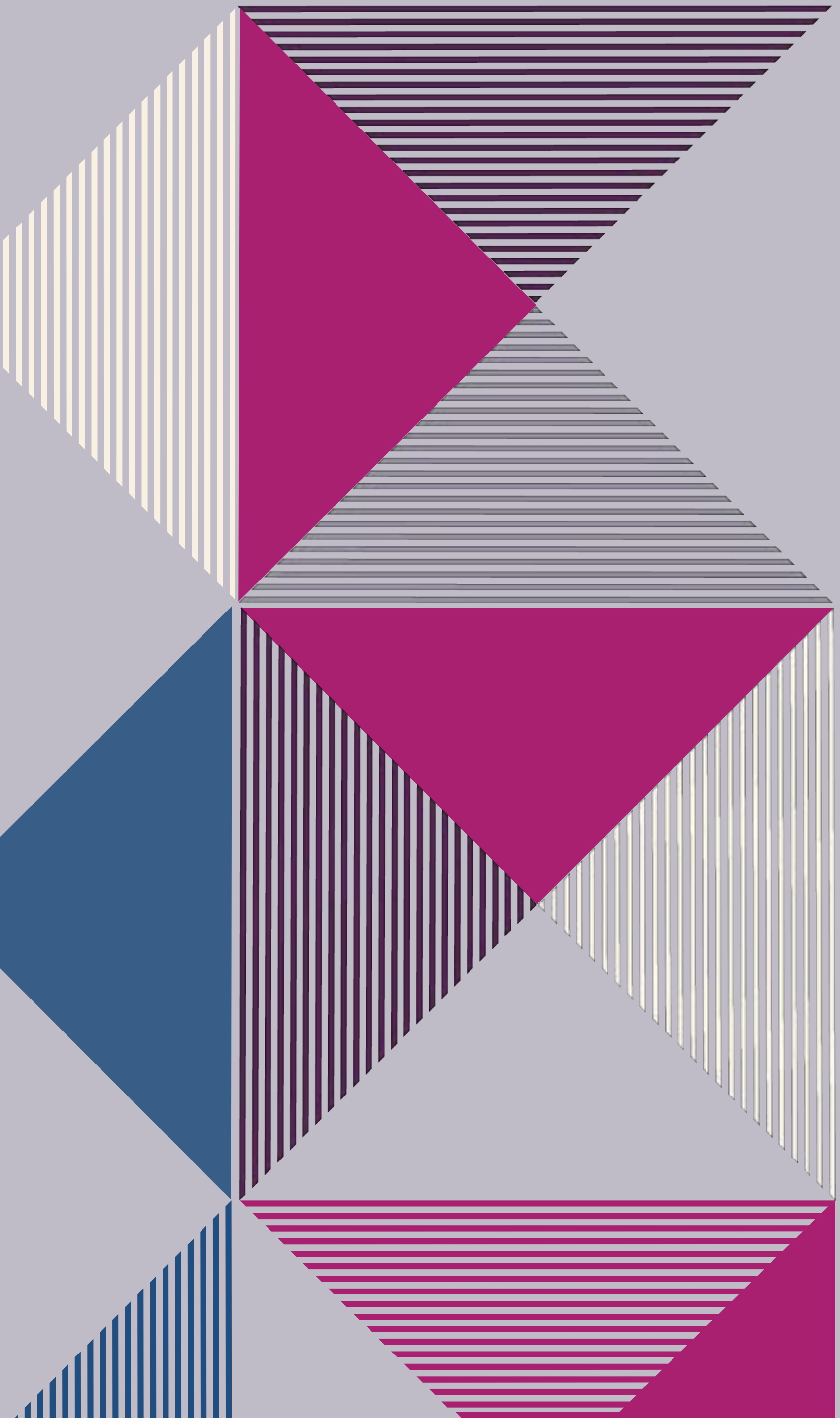
• The dark purple cluster has negative in profit and margin indicate at some points the stores didn’t do good--could be the first start opening.

SUMMARY

- Relationships: According to analysis, there is a medium positive linear relationship between the sales and profit. We will have to look at the total expenses and marketing as well because they have highest correlations.
- Regions: California(288), Colorado(264), Washington(240) and Ohio(214) are the highest sales.
- Product: Regular Expresso and Earl Grey are the highest sales and made profit.
- Market Type: The Major Market made higher sales , profit and COGs than the small market

Deliverables:
All analysis and suggestions have been collected in a Tableau Public

Need extra information?
Please click here and check my GitHub repository



INSTACART BASKET DATA ANALYSIS

OBJECTIVE

Provide an analysis of Instacart's sales patterns that will show customer behavior to help develop marketing and sales strategies to increase revenue

TOOLS & SKILLS

- Data wrangling and data frame merging in Python
- Deriving new variables
- Crosstabs and pivot tables in Python
- Visualizations in multiple Python libraries
- Markup and notebook management in Jupyter

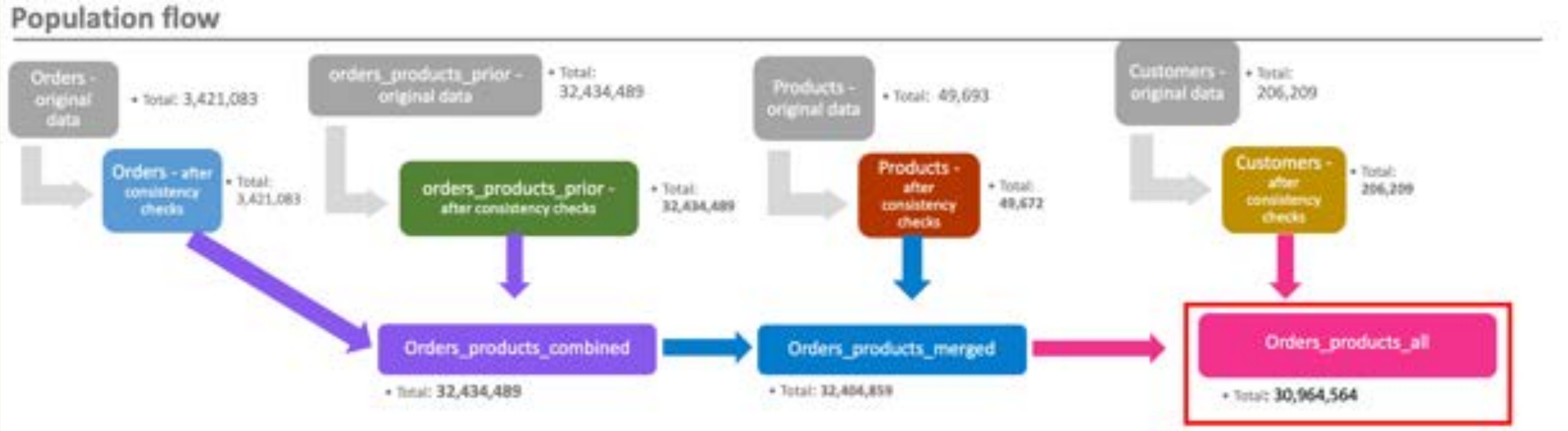
DATA

Open source data from Instacart and a customer data set created for the purpose of this project. Customers Data Set

LIMITATION

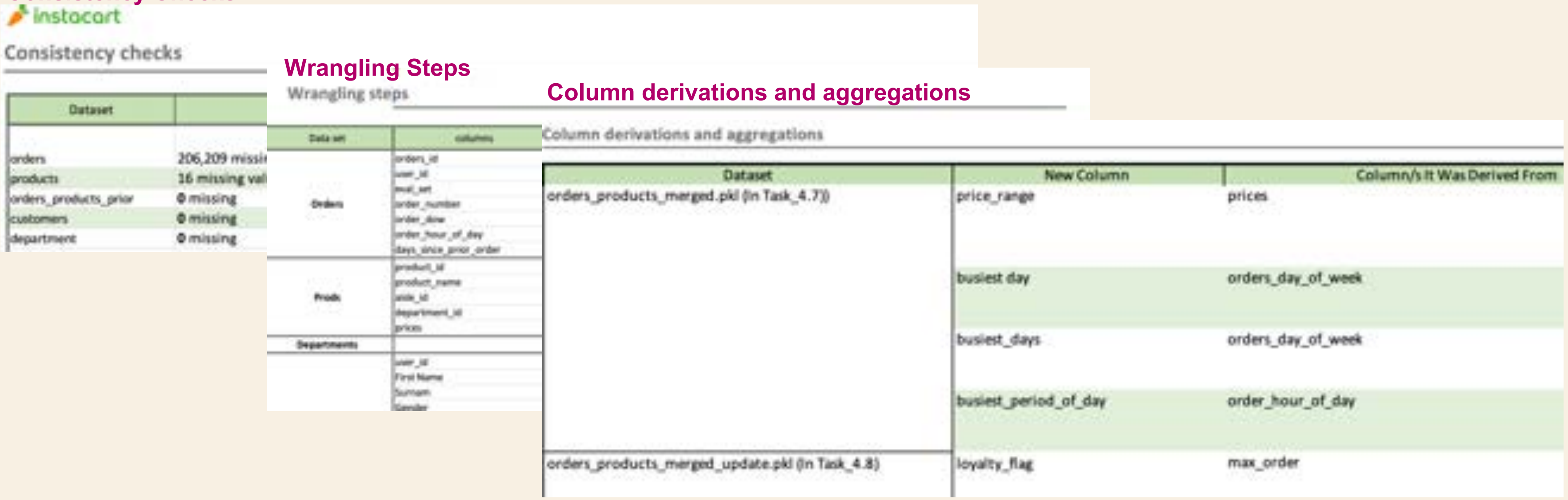
- Data only contains records from 2017
- Customer demographics are limited, only including age, family size, income, and marital status

ANALYSIS AND INSIGHT



Population Flow gives an overview of all merging phases. Different datasets have been merged to reach the most complete and up to date dataset.

Consistency Checks



instacart

Consistency checks

Dataset	
orders	206,209 missi
products	16 missing val
orders_products_prior	0 missing
customers	0 missing
department	0 missing

Wrangling Steps

Wrangling steps

Data set	columns
Orders	orders_id user_id eval_set order_number order_dow order_hour_of_day days_since_prior_order
Prods	product_id product_name aisle_id department_id prices
Departments	
	user_id First Name Surnam Gender

Column derivations and aggregations

Column derivations and aggregations

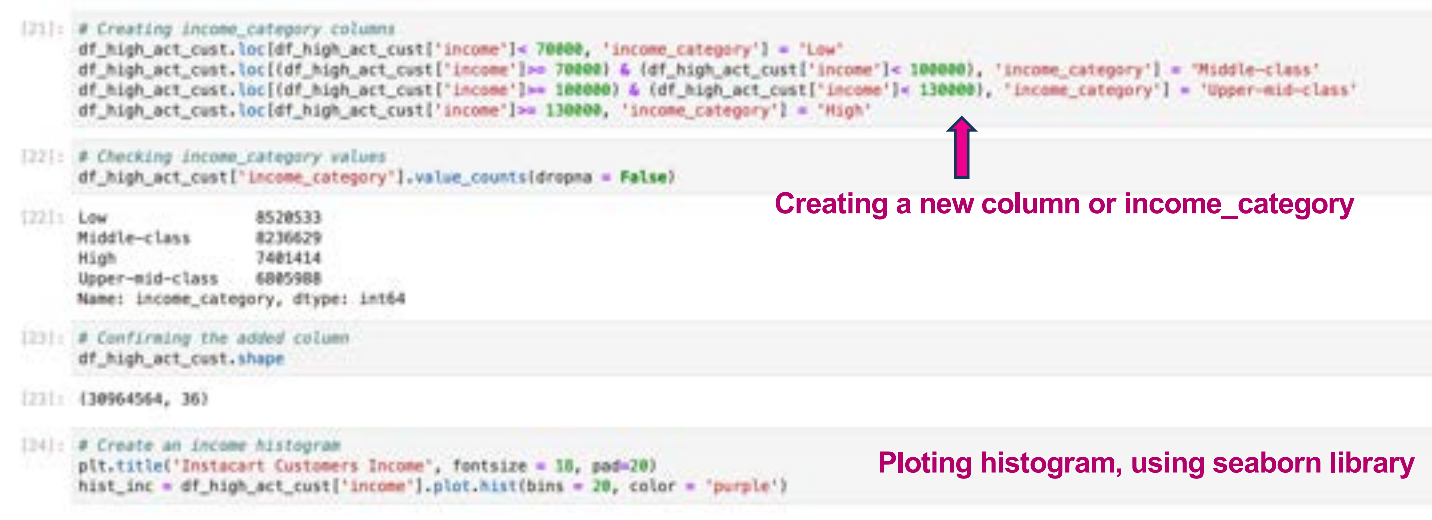
Dataset	New Column	Column/s It Was Derived From
orders_products_merged.pkl (in Task_4.7))	price_range	prices
	busiest day	orders_day_of_week
	busiest_days	orders_day_of_week
	busiest_period_of_day	order_hour_of_day
orders_products_merged_update.pkl (in Task_4.8)	loyalty_flag	max_order

Consistency Checks:
Checking if values are missing or duplicate and checking for mixed type variables.

Wrangling steps:
Changing columns headers and data types or creating new data frames.

Column derivations and aggregations
Creating new columns/variables and aggregated variables.

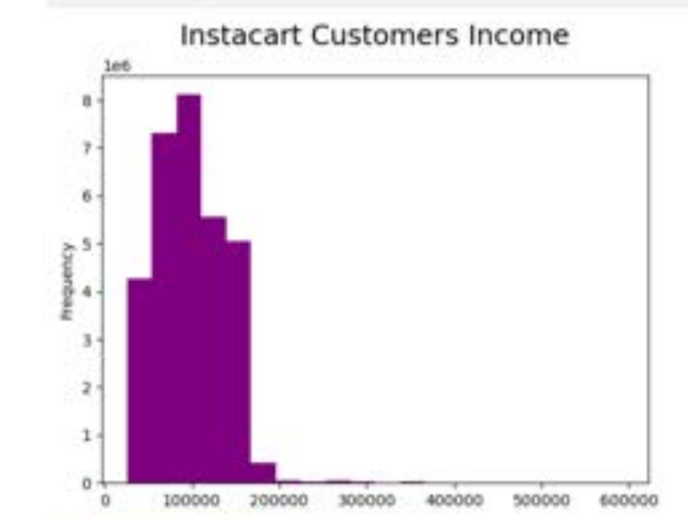
ANALYSIS AND INSIGHT



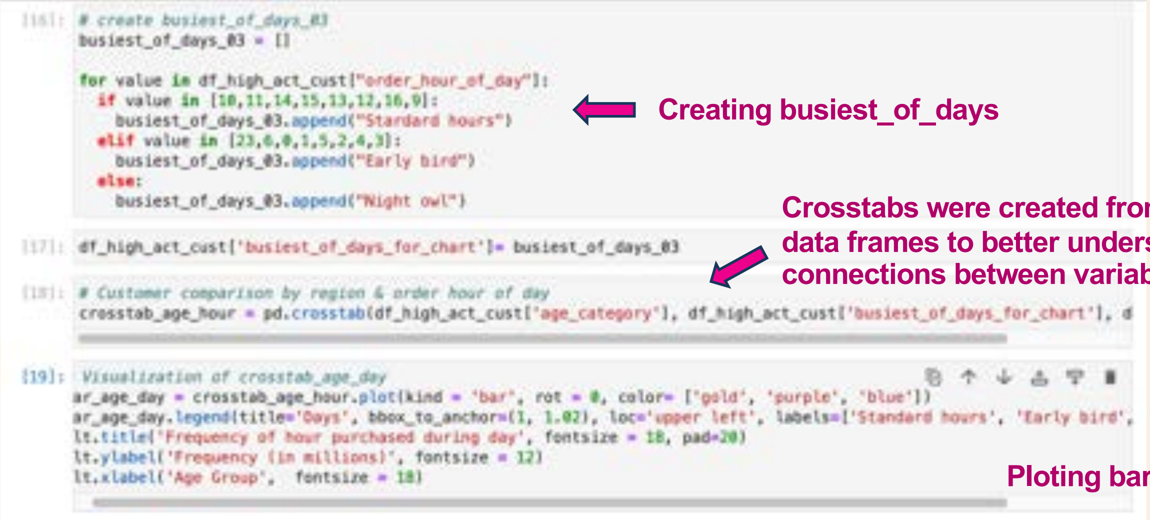
```
[21]: # Creating income_category columns
      df_high_act_cust.loc[df_high_act_cust['income']< 70000, 'income_category'] = 'Low'
      df_high_act_cust.loc[(df_high_act_cust['income']>= 70000) & (df_high_act_cust['income']< 100000), 'income_category'] = 'Middle-class'
      df_high_act_cust.loc[(df_high_act_cust['income']>= 100000) & (df_high_act_cust['income']< 130000), 'income_category'] = 'Upper-mid-class'
      df_high_act_cust.loc[df_high_act_cust['income']>= 130000, 'income_category'] = 'High'

[22]: # Checking income_category values
      df_high_act_cust['income_category'].value_counts(dropna = False)

[22]: Low                8520533
      Middle-class       8236629
      High               7401414
      Upper-mid-class    6805988
      Name: income_category, dtype: int64

[23]: # Confirming the added column
      df_high_act_cust.shape

[23]: (30964564, 36)

[24]: # Create an income histogram
      plt.title('Instacart Customers Income', fontsize = 18, pad=20)
      hist_inc = df_high_act_cust['income'].plot.hist(bins = 20, color = 'purple')
```

Creating a new column or income_category

Ploting histogram, using seaborn library



Customers with low income has the highest order

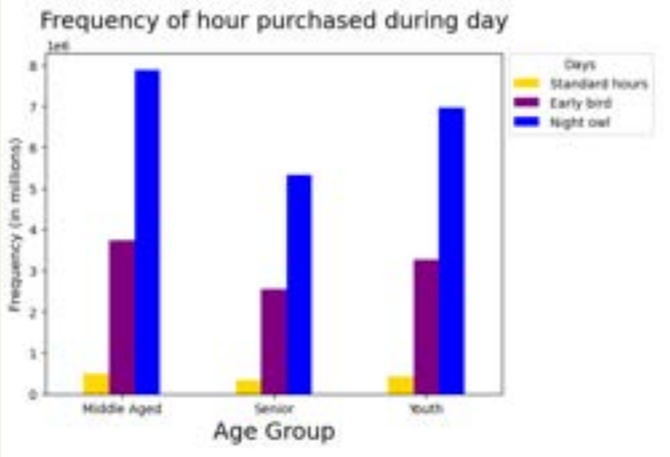


```
[16]: # create busiest_of_days_03
      busiest_of_days_03 = []

      for value in df_high_act_cust["order_hour_of_day"]:
        if value in [10,11,14,15,13,12,16,9]:
          busiest_of_days_03.append("Standard hours")
        elif value in [23,6,0,1,5,2,4,3]:
          busiest_of_days_03.append("Early bird")
        else:
          busiest_of_days_03.append("Night owl")

[17]: df_high_act_cust['busiest_of_days_for_chart']= busiest_of_days_03

[18]: # Customer comparison by region & order hour of day
      crosstab_age_hour = pd.crosstab(df_high_act_cust['age_category'], df_high_act_cust['busiest_of_days_for_chart'], d

[19]: Visualization of crosstab_age_day
     ar_age_day = crosstab_age_hour.plot(kind = 'bar', rot = 0, color= ['gold', 'purple', 'blue'])
     ar_age_day.legend(title='Days', bbox_to_anchor=(1, 1.02), loc='upper left', labels=['Standard hours', 'Early bird',
     lt.title('Frequency of hour purchased during day', fontsize = 18, pad=20)
     lt.ylabel('Frequency (in millions)', fontsize = 12)
     lt.xlabel('Age Group',  fontsize = 18)
```

Creating busiest_of_days

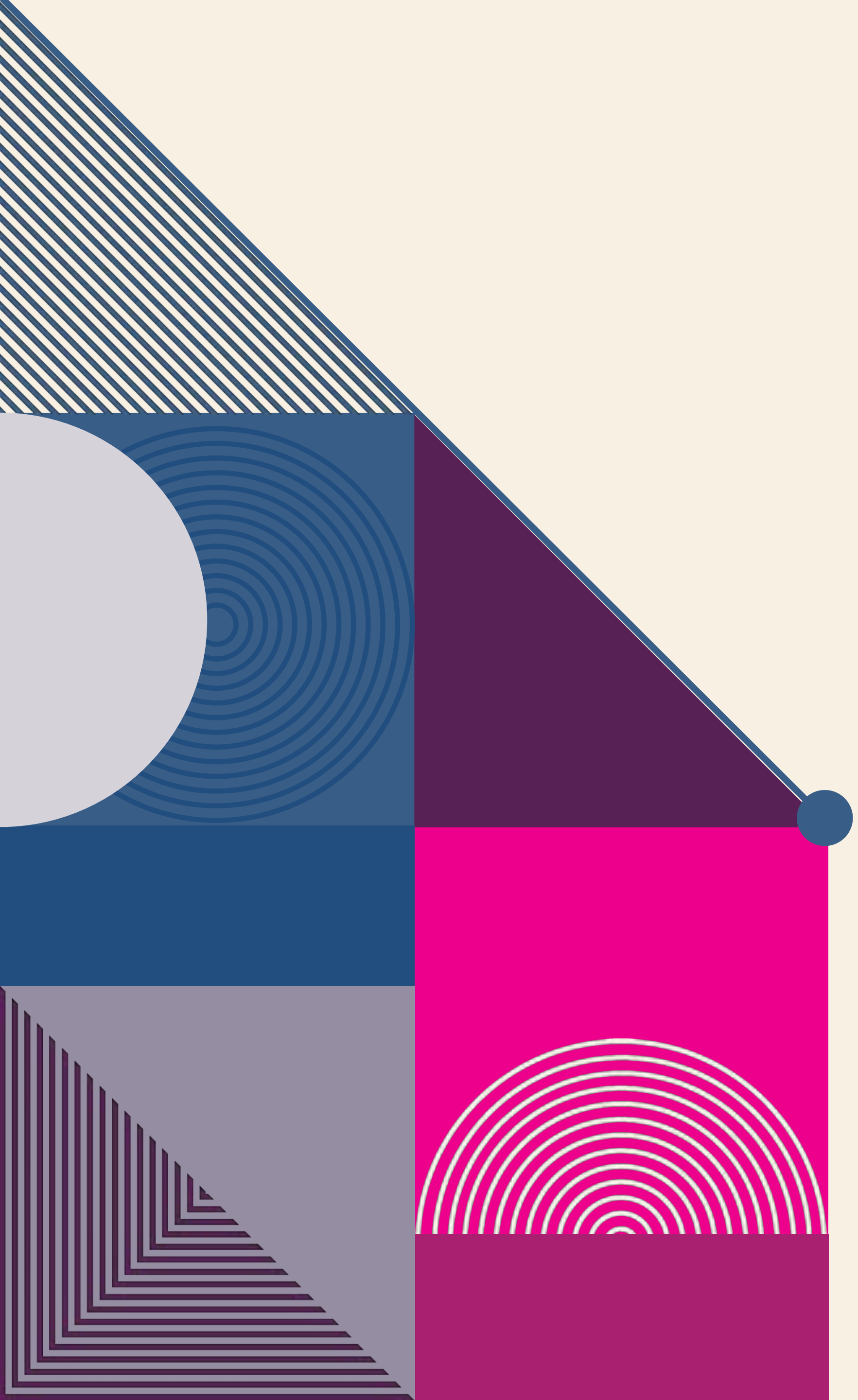
Crosstabs were created from the merged data frames to better understand the connections between variables.

Ploting barchart



Here is 5 Business key questions python code [Github](#)

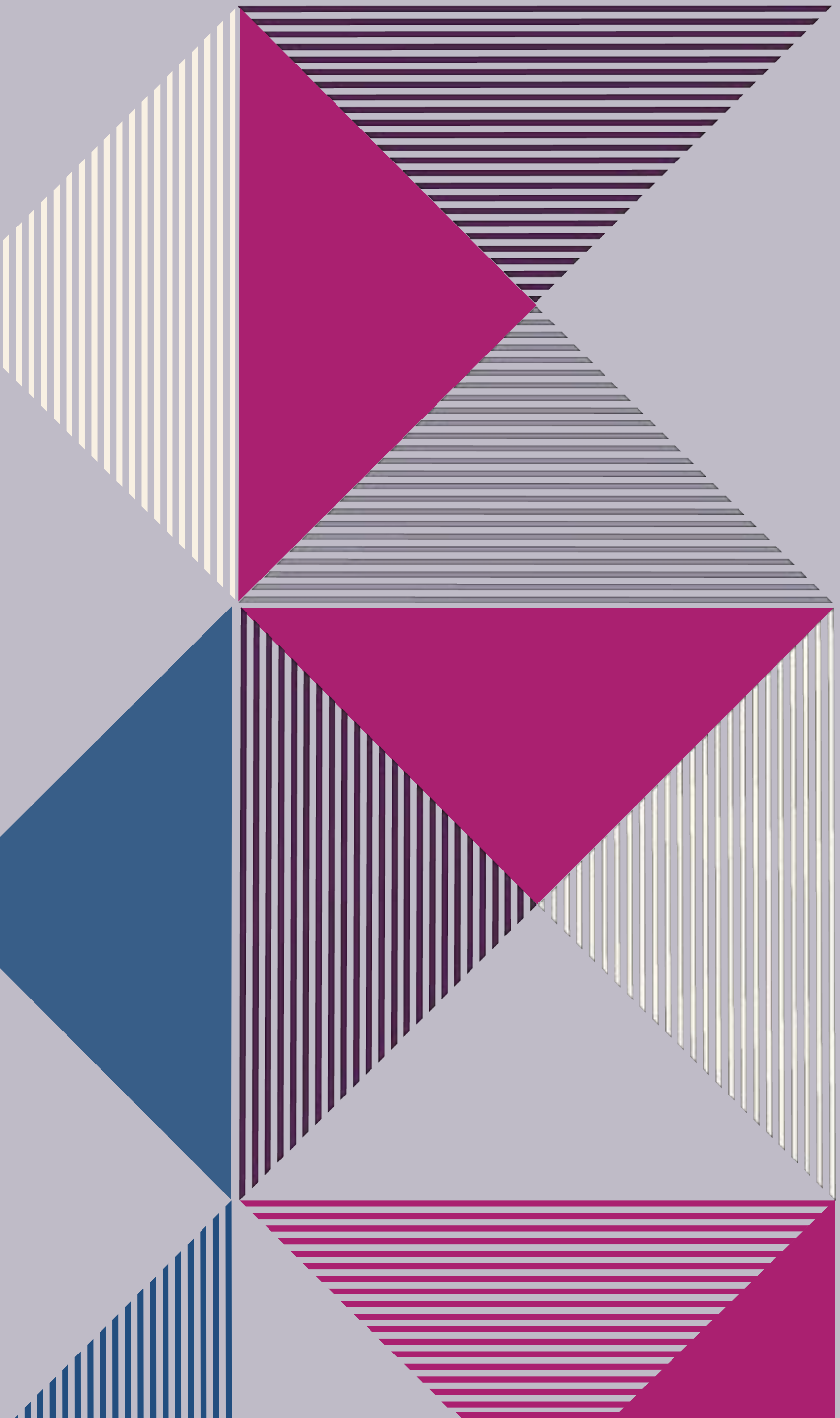
Middle-aged people tend to purchase the most items and at night Or after work.



RECOMMENDATIONS

- **Ads**–Schedule advertisements on the busy weekends so it reaches as many people as possible, specifically between 10am and 2pm.
- **Pricing**–Expand the market of higher priced items to boost their numbers and bring in more revenue.
- **Products**–The most popular products being ordered are those in produce, dairy and eggs, snacks, beverages. Instacart should carry on advertising those product and potentially offering deals to drive sales.
- **Loyalty**–Most customers are new or regular. To ensure that new customers continue to return so Instacart could consider giving percentage discounts for orders to new users to increase uptake.
- **Geography**-The Southern customers tend to be regular customers in terms of ordering-time habits, they also tend to fall into the low-income class. Using this region to test new products would be beneficial and we should focus on growing the customer bases in other regions.

Deliverables:
All analysis and suggestions have been collected in a [Tableau Public]() and check my [GitHub]() repository



ROCKBUSTER STEALTH LLC

OBJECTIVE

Rockbuster Stealth is a fictional movie rental company with stores over the world. They ‘re planning to launch an online video rental service in order to stay competitive

TOOLS & SKILLS

- Relational databases in SQL
- Entity relationship diagram creation and usage
- Data dictionary creation
- Database querying, filtering, and cleaning
- Joining tables in relational database
- Subqueries and common table expressions

DATA

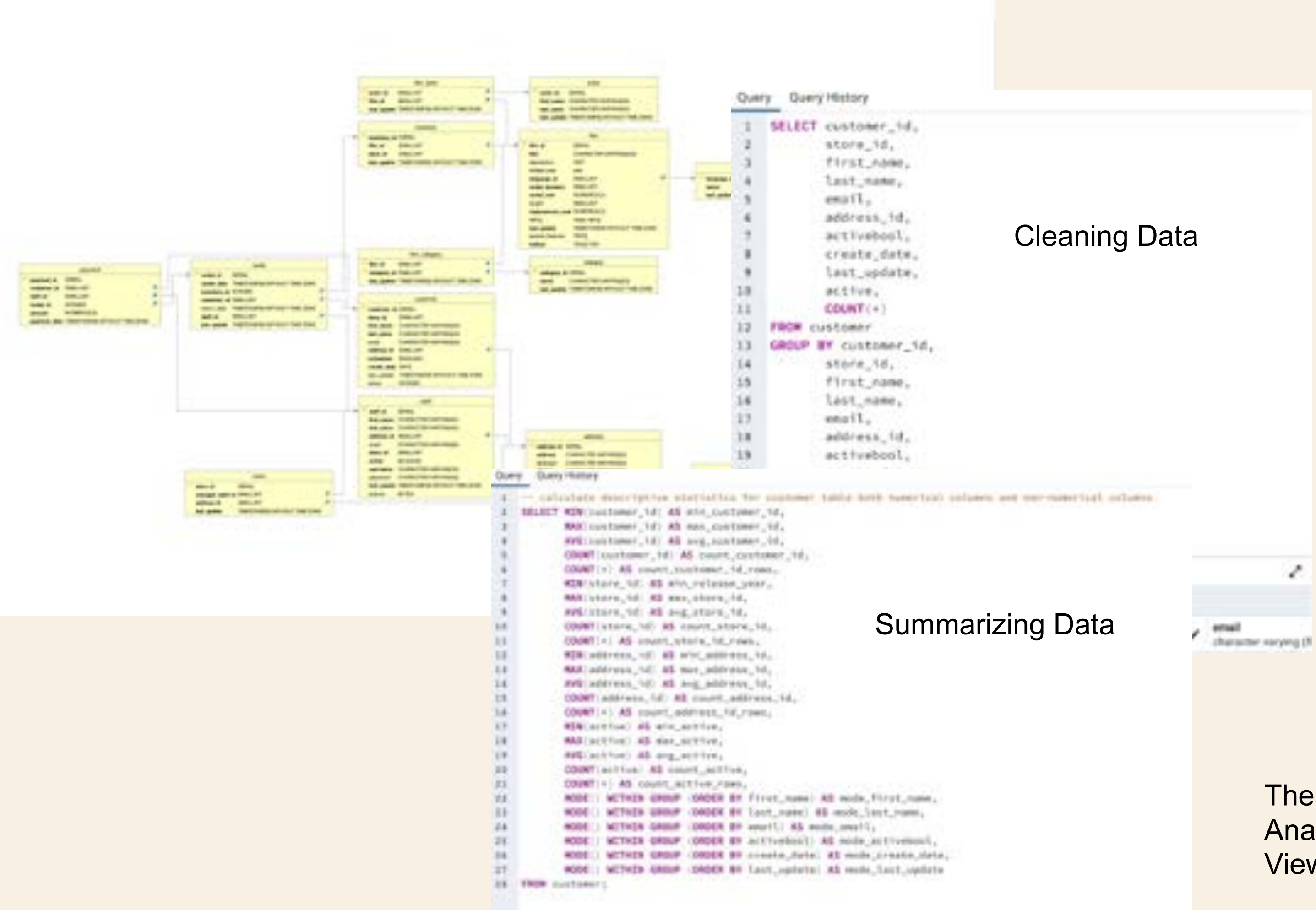
This dataset is provided by PostgreSQL for usage in tutorials. It contains data about film inventory, customers, payments, and associated details. Rockbuster Data Set

LIMITATION

- Only have internal records to work with provided by company

ANALYSIS AND INSIGHT

Entity Diagram



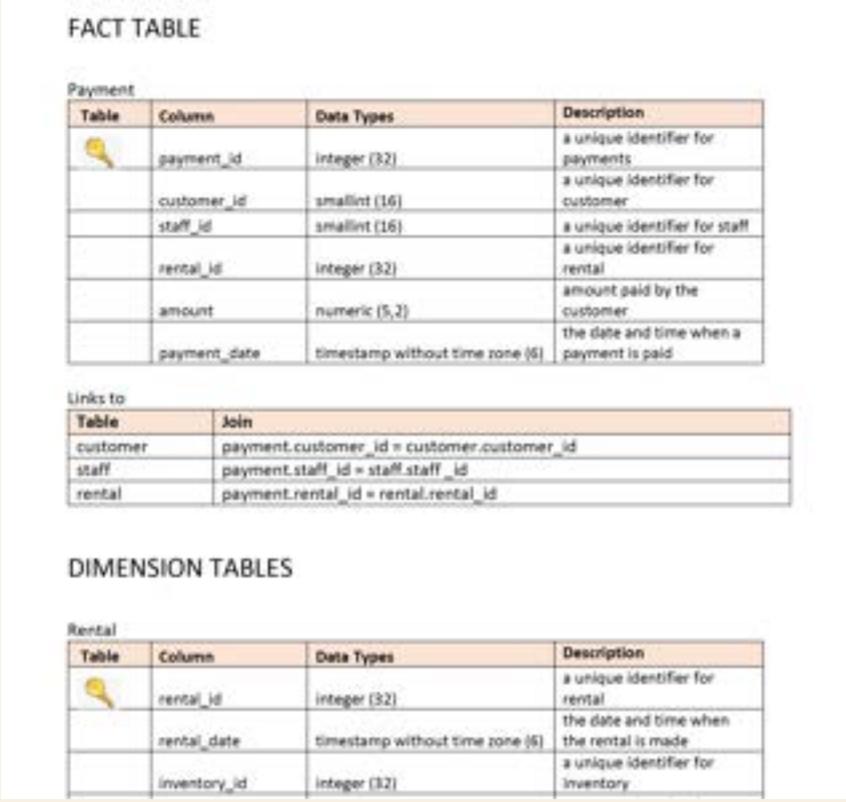
Cleaning Data

```sql
SELECT customer_id,
       store_id,
       first_name,
       last_name,
       email,
       address_id,
       activebool,
       create_date,
       last_update,
       active,
       COUNT(*)
FROM customer
GROUP BY customer_id,
         store_id,
         first_name,
         last_name,
         email,
         address_id,
         activebool,
```

Summarizing Data

```sql
-- calculate descriptive statistics for customer table both numerical columns and non-numerical columns
SELECT MIN(customer_id) AS min_customer_id,
       MAX(customer_id) AS max_customer_id,
       AVG(customer_id) AS avg_customer_id,
       COUNT(customer_id) AS count_customer_id,
       COUNT(*) AS count_customer_id_rows,
       MIN(store_id) AS min_release_year,
       MAX(store_id) AS max_store_id,
       AVG(store_id) AS avg_store_id,
       COUNT(store_id) AS count_store_id,
       COUNT(*) AS count_store_id_rows,
       MIN(address_id) AS min_address_id,
       MAX(address_id) AS max_address_id,
       AVG(address_id) AS avg_address_id,
       COUNT(address_id) AS count_address_id,
       COUNT(*) AS count_address_id_rows,
       MIN(active) AS min_active,
       MAX(active) AS max_active,
       AVG(active) AS avg_active,
       COUNT(active) AS count_active,
       COUNT(*) AS count_active_rows,
       MODE() WITHIN GROUP (ORDER BY first_name) AS mode_first_name,
       MODE() WITHIN GROUP (ORDER BY last_name) AS mode_last_name,
       MODE() WITHIN GROUP (ORDER BY email) AS mode_email,
       MODE() WITHIN GROUP (ORDER BY activebool) AS mode_activebool,
       MODE() WITHIN GROUP (ORDER BY create_date) AS mode_create_date,
       MODE() WITHIN GROUP (ORDER BY last_update) AS mode_last_update
FROM customer;
```

Data Dictionary



FACT TABLE

Payment

Table	Column	Data Types	Description
🔑	payment_id	integer (32)	a unique identifier for payments
	customer_id	smallint (16)	a unique identifier for customer
	staff_id	smallint (16)	a unique identifier for staff
	rental_id	integer (32)	a unique identifier for rental
	amount	numeric (5,2)	amount paid by the customer
	payment_date	timestamp without time zone (6)	the date and time when a payment is paid

Links to

Table	Join
customer	payment.customer_id = customer.customer_id
staff	payment.staff_id = staff.staff _id
rental	payment.rental_id = rental.rental_id

DIMENSION TABLES

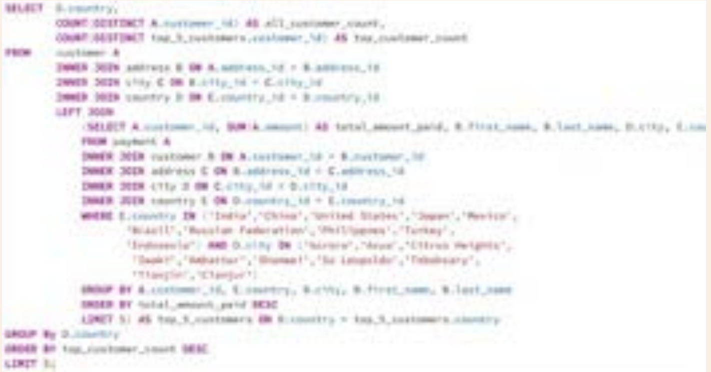
Rental

Table	Column	Data Types	Description
🔑	rental_id	integer (32)	a unique identifier for rental
	rental_date	timestamp without time zone (6)	the date and time when the rental is made
	inventory_id	integer (32)	a unique identifier for inventory

These are the process (Cleaning, Summarizing, Performing Descriptive Analysis) to help to perform query and analyze the data.
View [Data Dictionary](#)

ANALYSIS AND INSIGHT

Subquery

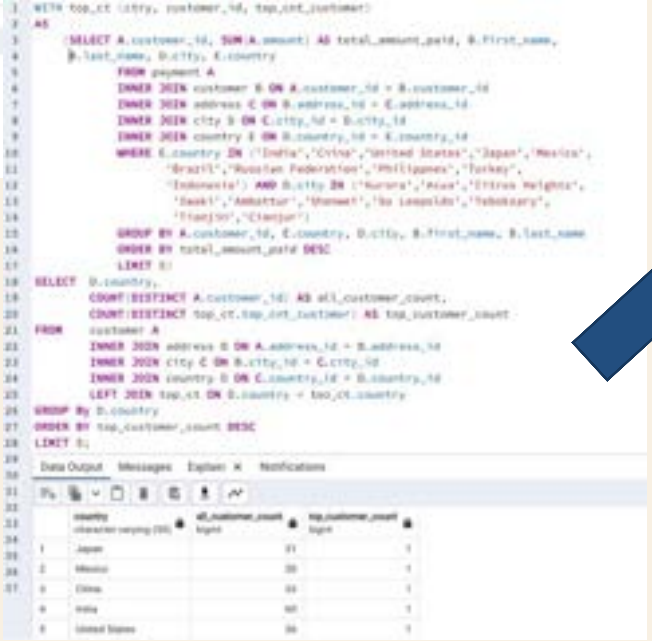


```
SELECT  D.country,
        COUNT(DISTINCT A.customer_id) AS all_customer_count,
        COUNT(DISTINCT top_5_customers.customer_id) AS top_customer_count
FROM    customer A
        INNER JOIN address B ON A.address_id = B.address_id
        INNER JOIN city C ON B.city_id = C.city_id
        INNER JOIN country D ON C.country_id = D.country_id
        LEFT JOIN
            (SELECT A.customer_id, SUM(A.amount) AS total_amount_paid, B.first_name, B.last_name, D.city, E.cou
            FROM payment A
            INNER JOIN customer B ON A.customer_id = B.customer_id
            INNER JOIN address C ON B.address_id = C.address_id
            INNER JOIN city D ON C.city_id = D.city_id
            INNER JOIN country E ON D.country_id = E.country_id
            WHERE E.country IN ('India','China','United States','Japan','Mexico',
                  'Brazil','Russian Federation','Philippines','Turkey',
                  'Indonesia') AND D.city IN ('Aurora','Acua','Citrus Heights',
                   'Iwaki','Ambattur','Shanwei','So Leopoldo','Teboksary',
                   'Tianjin','Cianjur')
            GROUP BY A.customer_id, E.country, D.city, B.first_name, B.last_name
            ORDER BY total_amount_paid DESC
            LIMIT 5) AS top_5_customers ON D.country = top_5_customers.country
GROUP By D.country
ORDER BY top_customer_count DESC
LIMIT 5;
```

Data Output | Messages | Notifications

	country character varying (50)	all_customer_count bigint	top_customer_count bigint
1	Japan	31	1
2	Mexico	30	1
3	China	53	1
4	India	60	1
5	United States	36	1

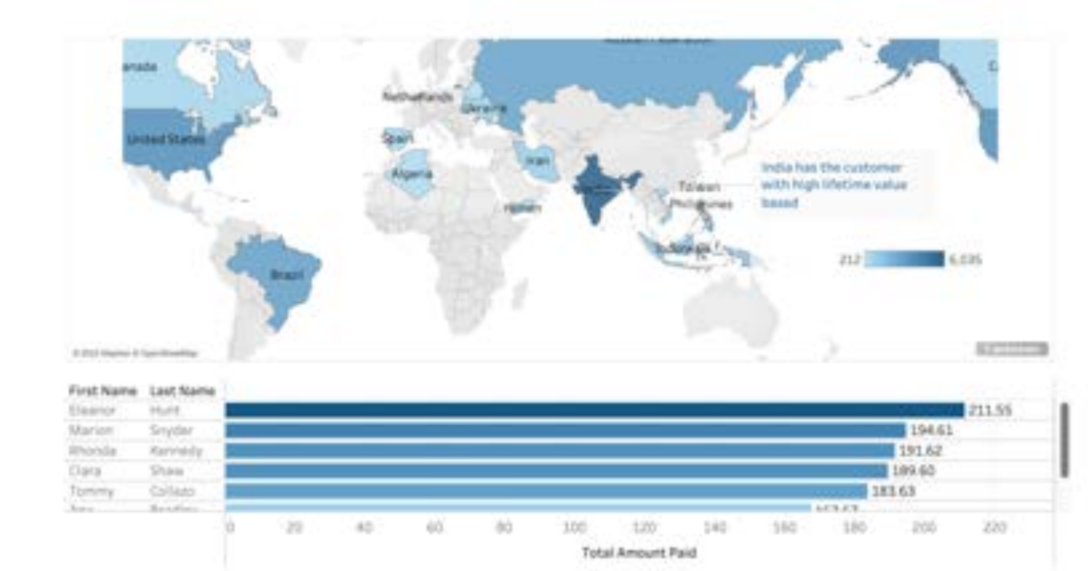
CTE



```
WITH top_ct (city, customer_id, top_cnt_customer)
AS
    (SELECT A.customer_id, SUM(A.amount) AS total_amount_paid, B.first_name,
    B.last_name, D.city, E.country
            FROM payment A
            INNER JOIN customer B ON A.customer_id = B.customer_id
            INNER JOIN address C ON B.address_id = C.address_id
            INNER JOIN city D ON C.city_id = D.city_id
            INNER JOIN country E ON D.country_id = E.country_id
            WHERE E.country IN ('India','China','United States','Japan','Mexico',
                  'Brazil','Russian Federation','Philippines','Turkey',
                  'Indonesia') AND D.city IN ('Aurora','Acua','Citrus Heights',
                   'Iwaki','Ambattur','Shanwei','So Leopoldo','Teboksary',
                   'Tianjin','Cianjur')
            GROUP BY A.customer_id, E.country, D.city, B.first_name, B.last_name
            ORDER BY total_amount_paid DESC
            LIMIT 5)
SELECT  D.country,
        COUNT(DISTINCT A.customer_id) AS all_customer_count,
        COUNT(DISTINCT top_ct.top_cnt_customer) AS top_customer_count
FROM    customer A
        INNER JOIN address B ON A.address_id = B.address_id
        INNER JOIN city C ON B.city_id = C.city_id
        INNER JOIN country D ON C.country_id = D.country_id
        LEFT JOIN top_ct ON D.country = top_ct.country
GROUP By D.country
ORDER BY top_customer_count DESC
LIMIT 5;
```

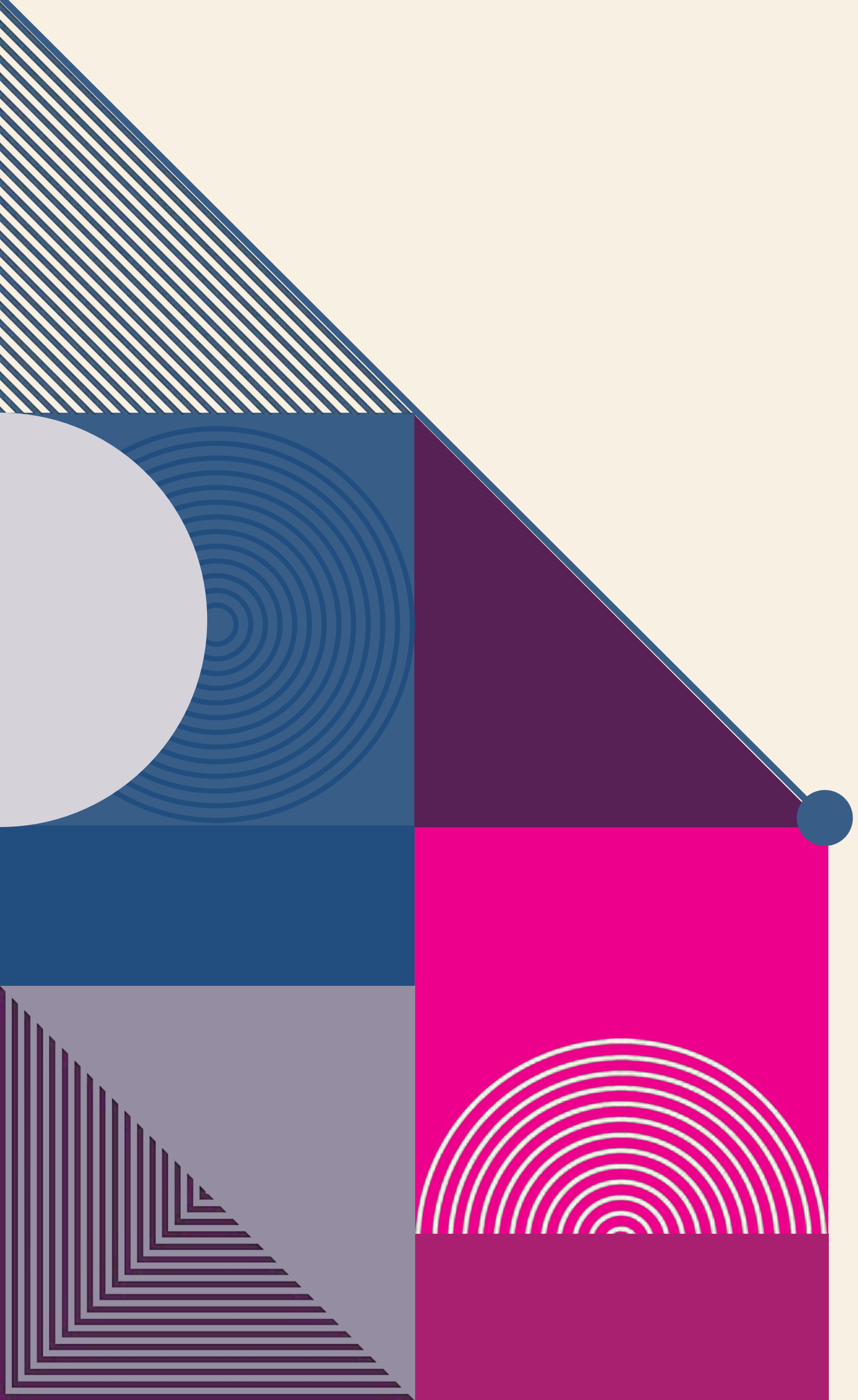
Data Output | Messages | Explain ✕ | Notifications

	country character varying (50)	all_customer_count bigint	top_customer_count bigint
1	Japan	31	1
2	Mexico	30	1
3	China	53	1
4	India	60	1
5	United States	36	1



To answer the business questions, the right table joins and queries had to be written in SQL. Then the resulting table was exported to a csv file and imported into Tableau. At that point, a visualization showing the answers to the business questions could be created. When perform complex queries, CTE is easier to organize and read. However, readability is not the only consideration when choosing between a CTE and a subquery; performance is also important

Here is [5 key business questions queries]



RECOMMENDATIONS

Revenue: To maximize revenue and customer satisfaction. Rockbuster should prioritize launching the top 10 highest movies on their online video rental service.

Customer: To expand their customer, Rockbuster should conduct further analysis to understand why other regions are not performing as well as Asia and America.

Location: To get a large and diverse customer base, Rockbuster should prioritize the Asia and American markets as their primary targets,

Deliverables:

All analysis and suggestions have been collected in a Tableau Public

Need extra information?

Please click here and check my GitHub repository

PREPARING FOR UP COMING FLU SEASON

OBJECTIVE

Identify geographic and seasonal trends for annual influenza outbreaks in the USA. Provide tools for a medical staffing agency to identify where and when to allocate additional medical support.

DATA

- CDC Influenzadeaths
- Population data by geography (US Census Bureau)

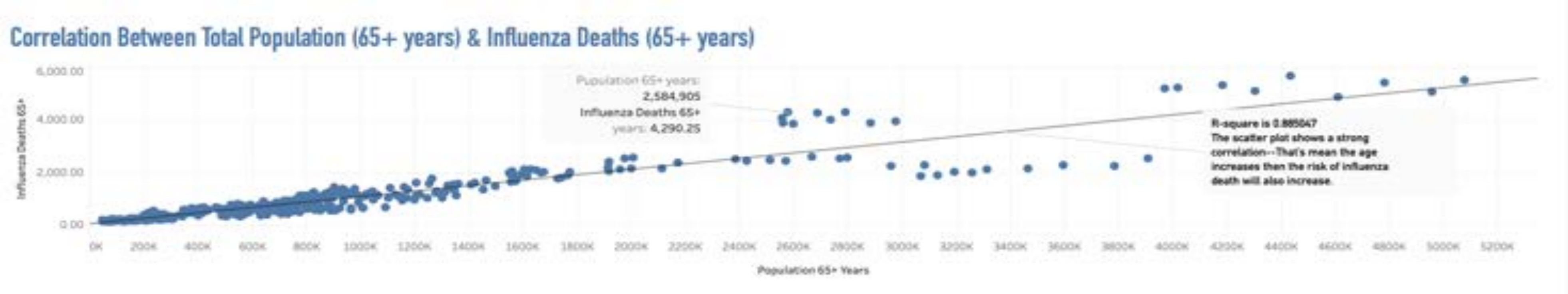
TOOLS & SKILLS

- Data research project design
- Data profiling and cleaning
- Data integration and transformation
- Statistical hypothesis testing
- Geographic visualizations and time-series forecasting
- Interactive visualizations and storytelling in Tableau

LIMITATION

- Data is dated from 2009 until 2017
- Influenza death data is 82% suppressed due to confidentiality
- Staffing capacity and sizes of hospitals/clinics is unknown

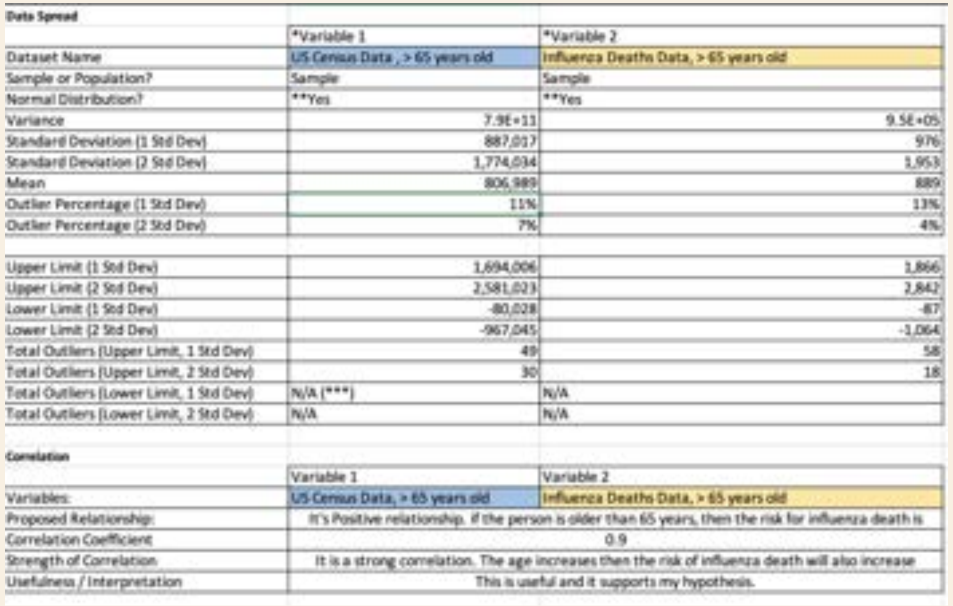
ANALYSIS AND INSIGHT



Correlation Between Total Population (65+ years) & Influenza Deaths (65+ years)

Population 65+ years: 2,584,905
Influenza Deaths 65+ years: 4,290.25

R-square is 0.885047
The scatter plot shows a strong correlation--That's mean the age increases then the risk of influenza death will also increase.



Data Spread

	*Variable 1	*Variable 2
Dataset Name	US Census Data , > 65 years old	Influenza Deaths Data, > 65 years old
Sample or Population?	Sample	Sample
Normal Distribution?	**Yes	**Yes
Variance	7.9E+11	9.5E+05
Standard Deviation (1 Std Dev)	887,017	976
Standard Deviation (2 Std Dev)	1,774,034	1,953
Mean	806,989	889
Outlier Percentage (1 Std Dev)	11%	13%
Outlier Percentage (2 Std Dev)	7%	4%
Upper Limit (1 Std Dev)	1,694,006	1,866
Upper Limit (2 Std Dev)	2,581,023	2,842
Lower Limit (1 Std Dev)	-80,028	-87
Lower Limit (2 Std Dev)	-967,045	-1,064
Total Outliers (Upper Limit, 1 Std Dev)	49	58
Total Outliers (Upper Limit, 2 Std Dev)	30	18
Total Outliers (Lower Limit, 1 Std Dev)	N/A (***)	N/A
Total Outliers (Lower Limit, 2 Std Dev)	N/A	N/A

Correlation

	Variable 1	Variable 2
Variables:	US Census Data, > 65 years old	Influenza Deaths Data, > 65 years old
Proposed Relationship:	It's Positive relationship. If the person is older than 65 years, then the risk for influenza death is	
Correlation Coefficient	0.9	
Strength of Correlation	It is a strong correlation. The age increases then the risk of influenza death will also increase	
Usefulness / Interpretation	This is useful and it supports my hypothesis.	

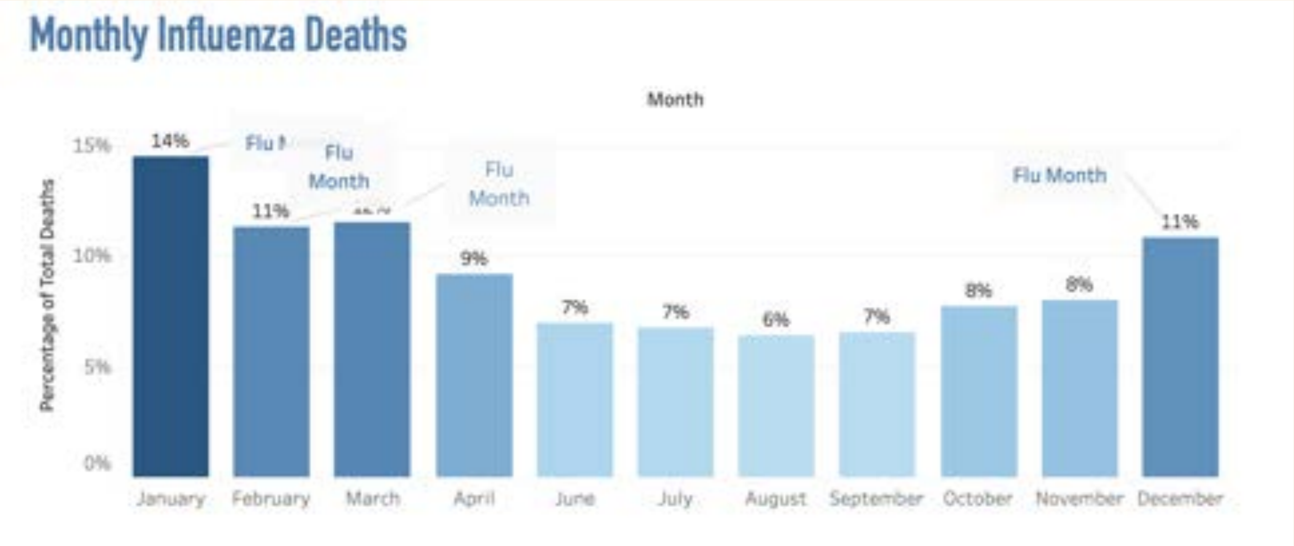
Research hypothesis

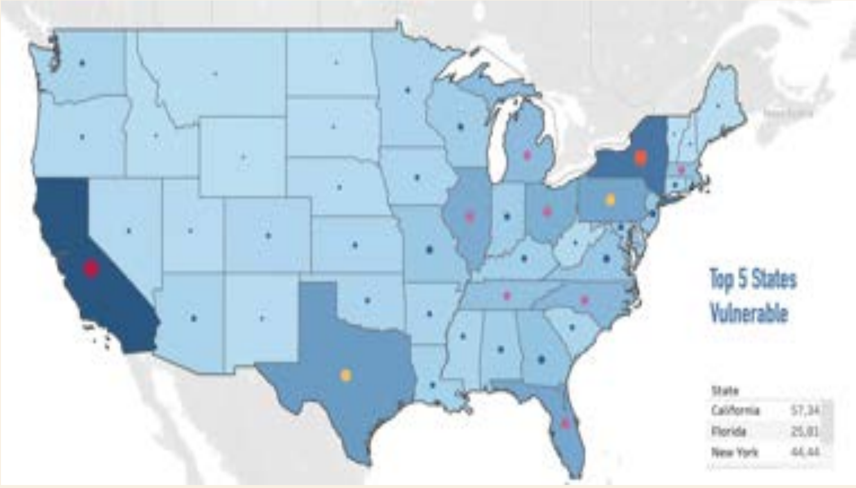
If persons who over the aged of 65 years, they are a higher risk for influenza death

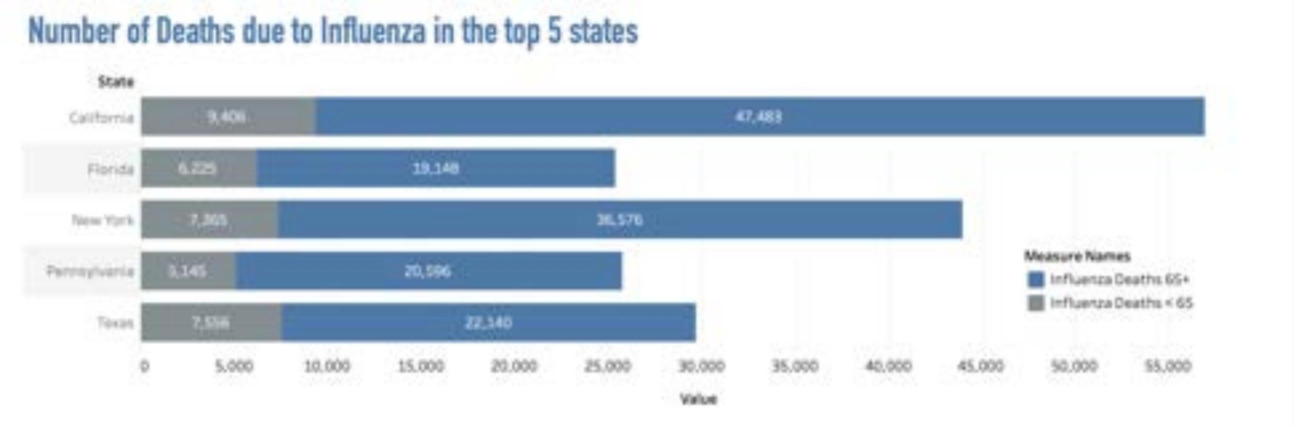
Descriptive Analysis

According to our hypothesis, mortality rate increases or is higher with increased age. The correlation study suggests strong correlation between age and mortality rate. The statistics for the same are summarized in table

ANALYSIS AND INSIGHT







Influenza Seasonal Peak:

December, January, February and March

The Most Populous States that have the Highest Number of Deaths:

California, New York, Texas, Pennsylvania, Florida. Ages of vulnerable population are 65 years and older

RECOMMENDATIONS

- **Staff:** Since the peak of flu is usually between December to January. More staff should be made available in this peak period. Staff deployments should also consider and prioritize the states that have the most flu deaths plus the prioritize of the vulnerable population (65 years and older)
- **Flu Shot:** Vaccines should be ready mostly in states with the most vulnerable population
- **Survey Evaluation:** To make sure your process is effective and adjust if needed. Recommend to monitor the effectiveness by using KPIs on Jan, Feb, March and April.

Deliverables:

All analysis and suggestions have been collected in a Tableau Public

GAMECO MARKETING DATA ANALYSIS

OBJECTIVE

Develop a current understanding of the global retail videogame sales market, to inform GameCo’s efforts to increase market share.

TOOLS & SKILLS

- Data quality, integrity, and consistency checks
- Data cleaning
- Pivot tables (data grouping & summarizing)
- Descriptive analysis
- Excel visualizations
- Reporting in PowerPoint

DATA

The data is made publicly available by VGChartz. It covers historical retail sales of videogames for games that sold more than 100,000 copies, until 2016.

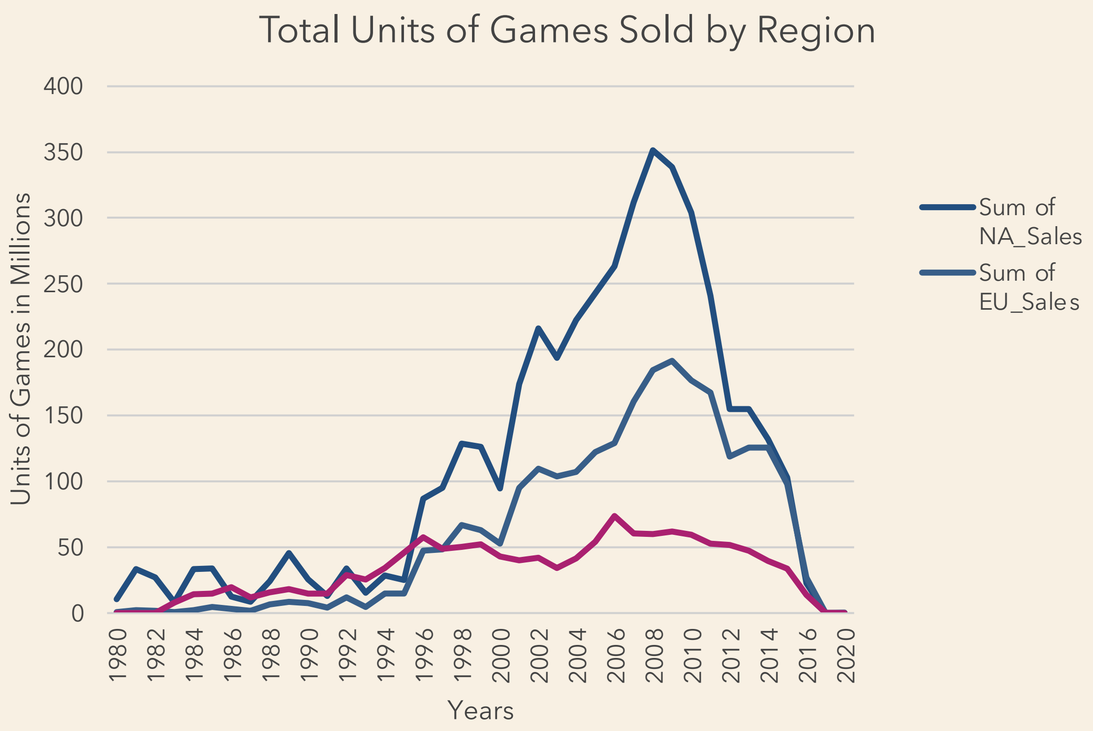
LIMITATION

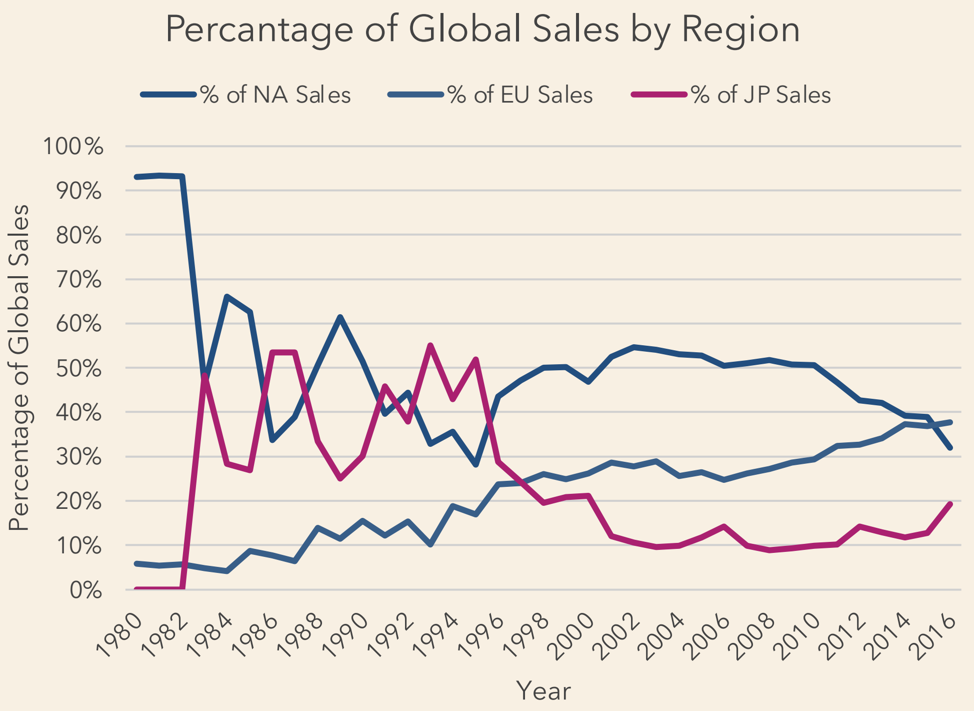
- The data only goes until 2016
- No revenue data, just units sold

ANALYSIS AND INSIGHT

- Using line chart that represent the proportion of global sales for North America, Europe and Japan by years. There are significant changes between them and should focus deeper insight.

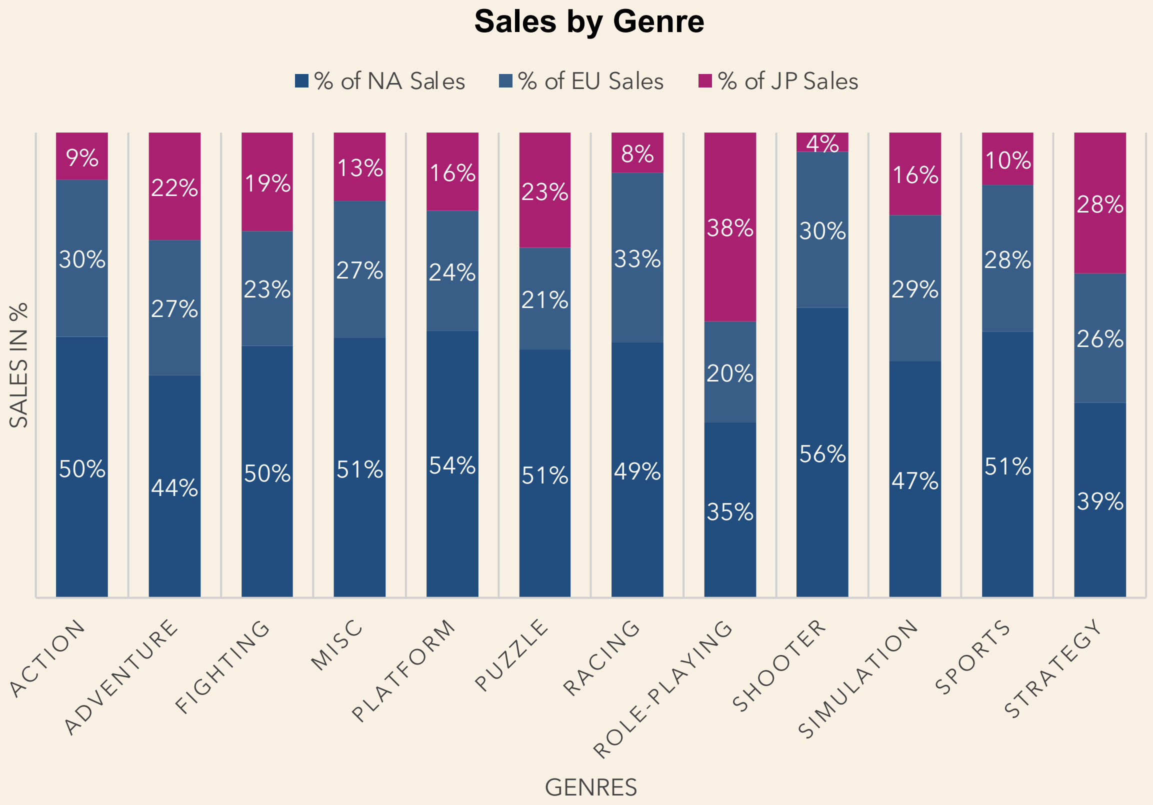
- After seeing the data behaves in 2016, GemeCo marketing team should investigate more in each region for seeing the deeper insights and see which region they should spend more marketing budget in 2017

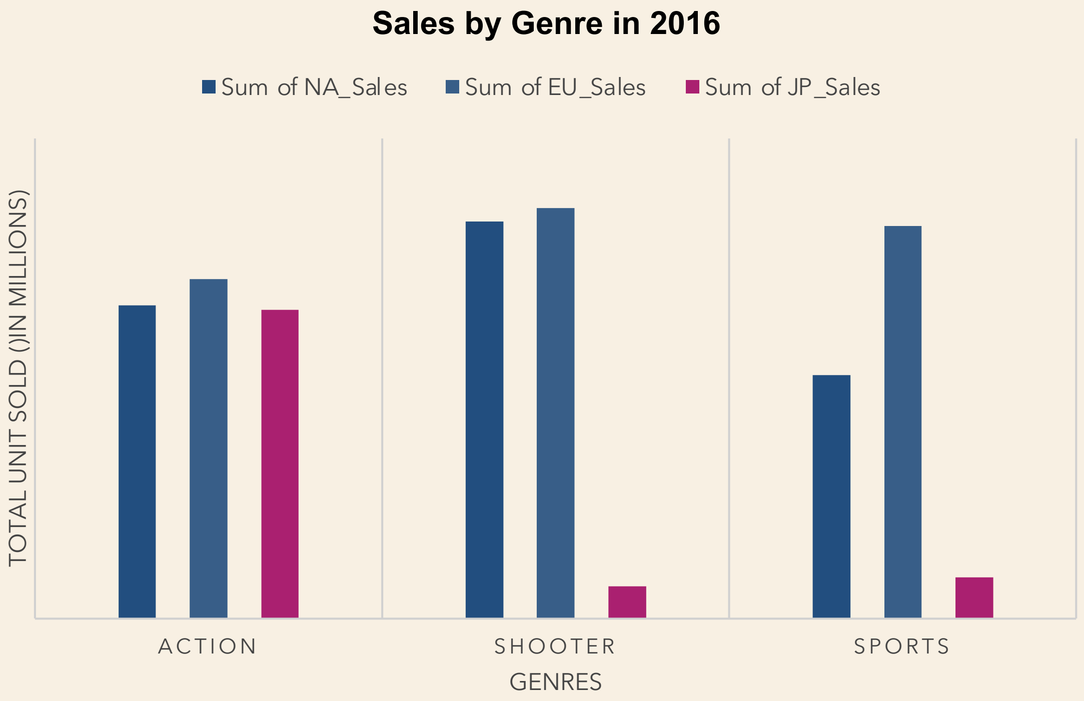


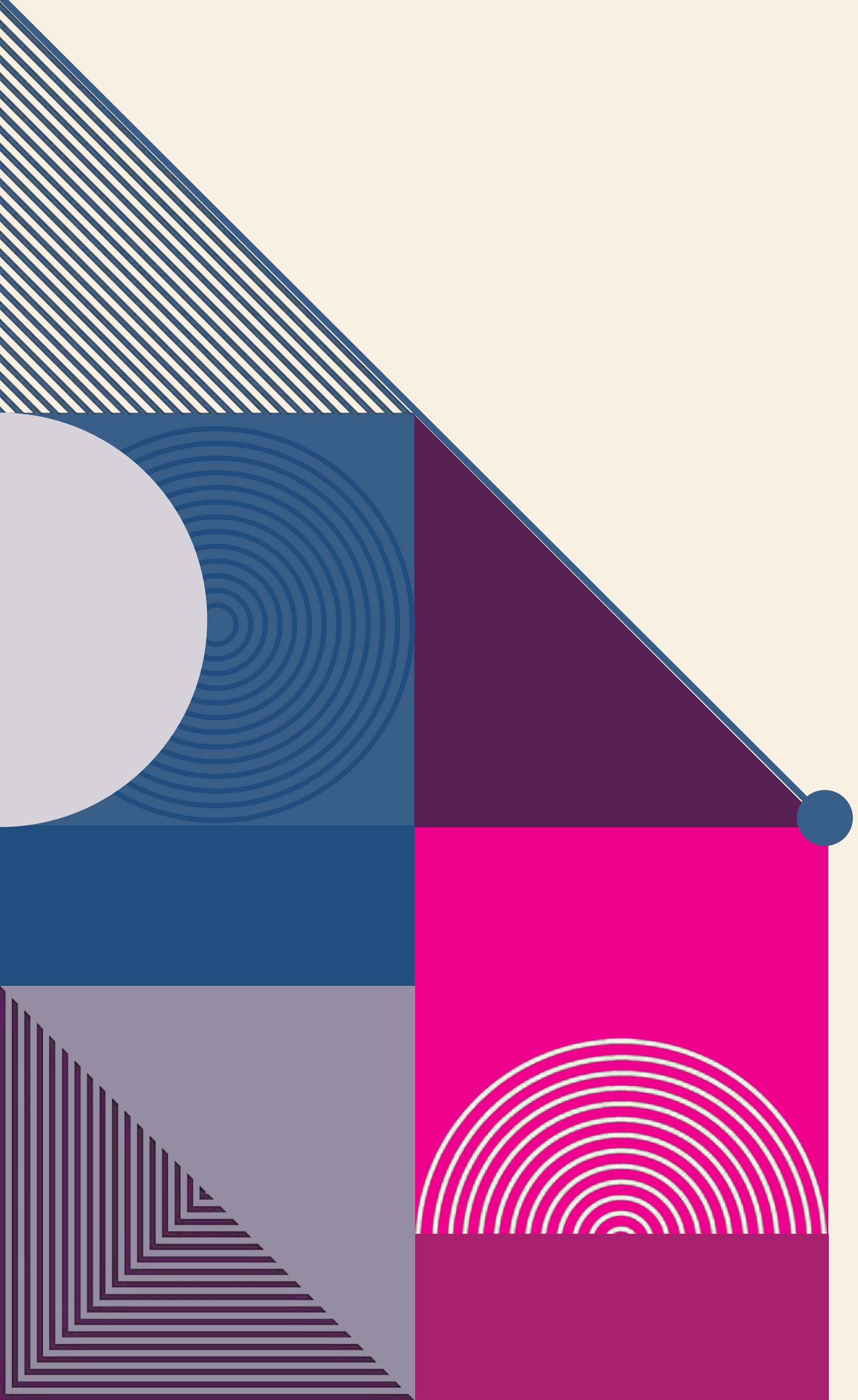


ANALYSIS AND INSIGHT

- Shooter games is the highest sales in genres for North America both overall and in the past year.
- Shooter game is the highest sales in genres for Europe over year however Action game also made the highest sales in genre when it represents the proportion of global sales over years
- Role-Playing game is the highest proportion for Japan sales over years however Action game has been in the top spot in 2016.







RECOMMENDATIONS

- **Budget:** Break it proportionally by region for sales numbers. Focus for the top selling genres within each region.
- **Marketing:** Put more money into the growing markets to increase revenue where demand is higher.
- **Growth**: North America is a huge market and sales there have been declining. Update and create a lot of new games (genres, title, publisher, platform) to help increase revenue there.

Deliverables:

All analysis and suggestions have been collected in a Tableau Public

THANK YOU

Rusamijan(Mee) Permison
meeps.analyst@gmail.com
Portfolio: www.meepermison.com