

Portfolio for Projects

Shi, Dongmei

17.04.2023

Outline of projects

Projects	Objectives	Tolls
<u>Video game popularity data analysis</u>	Analyzing data to inform the development of new games for a game company 'GameCo'	Excell
<u>Preparing for the influenza season</u>	Determine when to send staff, and how many to each state in USA	Excell, Tableau
<u>Rockbuster Stealth</u>	use data to help with the launch strategy for the new online video service	SQL, Tableau
<u>Instacart Grocery</u>	To uncover more information about the sales patterns	Python, Tableau

Project: Video Game Popularity

- Objective:
Analyzing data to inform the development of new games for a game company 'GameCo'
- Data sets:
'video_game_sales' data set that was drawn from VGChartz
- Tools:
Excel – Grouping, Summarizing and Descriptive analysis

The project report link:

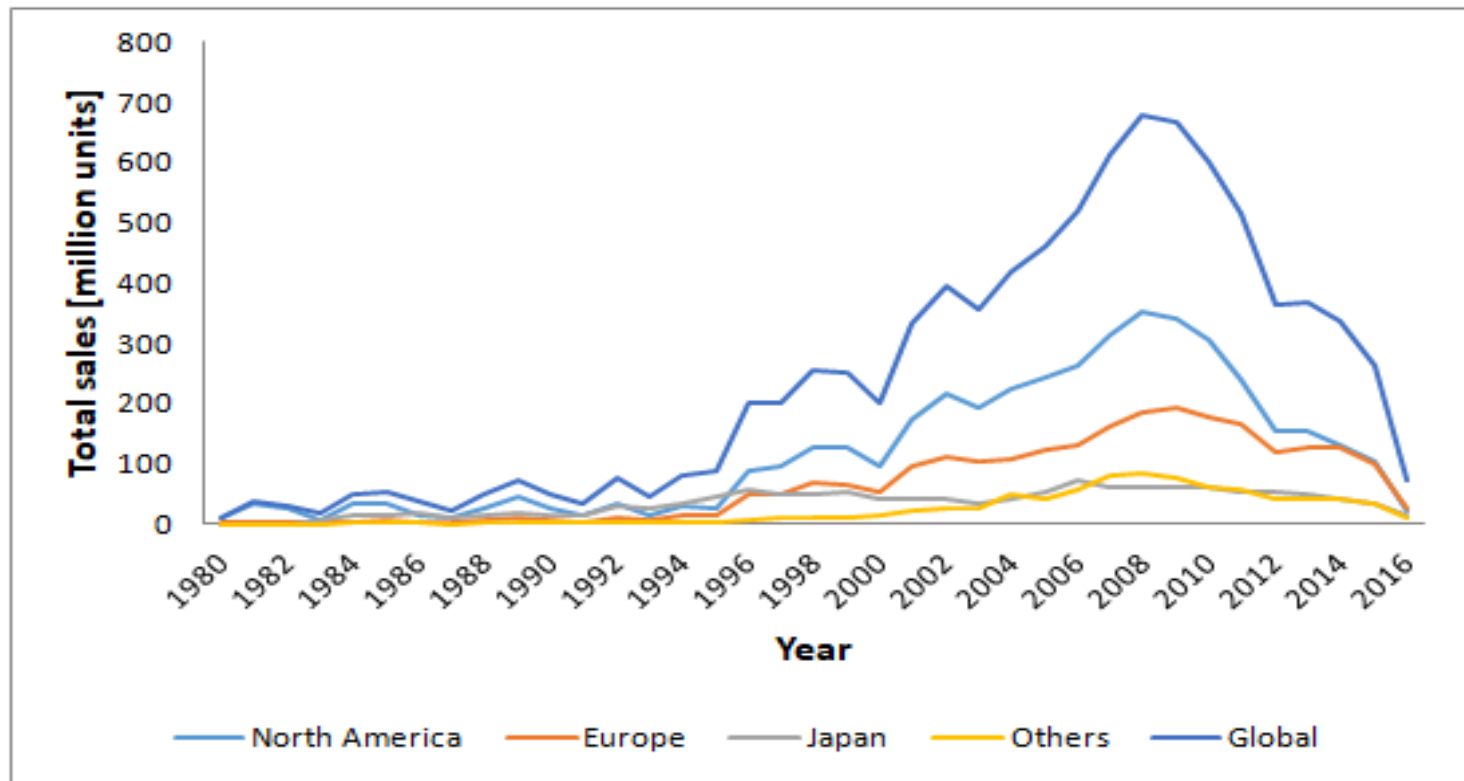
[Project-video-game-/Project- video game popularity.pdf at main · dongmei-s/Project-video-game- \(github.com\)](https://github.com/dongmei-s/Project-video-game-/blob/main/Project-video%20game%20popularity.pdf)

Analysis stages

Stages	Contents	Purpose
Verify the quality and integrity of data	Collect data, clean data, verify the completeness, accuracy and validity of data	To ensure the accuracy of the following analysis
Data analysis	Sales vs. time for each region, Sales distribution for regions, Sales distribution for genres.	To understand game sales vs. time regions and etc.
Modeling	Visualization by line chart and bar chart	To have a better understanding and make predictions
Conclusion and recommendations	Get conclusion and give recommendations	To answer the key questions the stakeholders care

Key visualizations - 1

- Sale's change over time

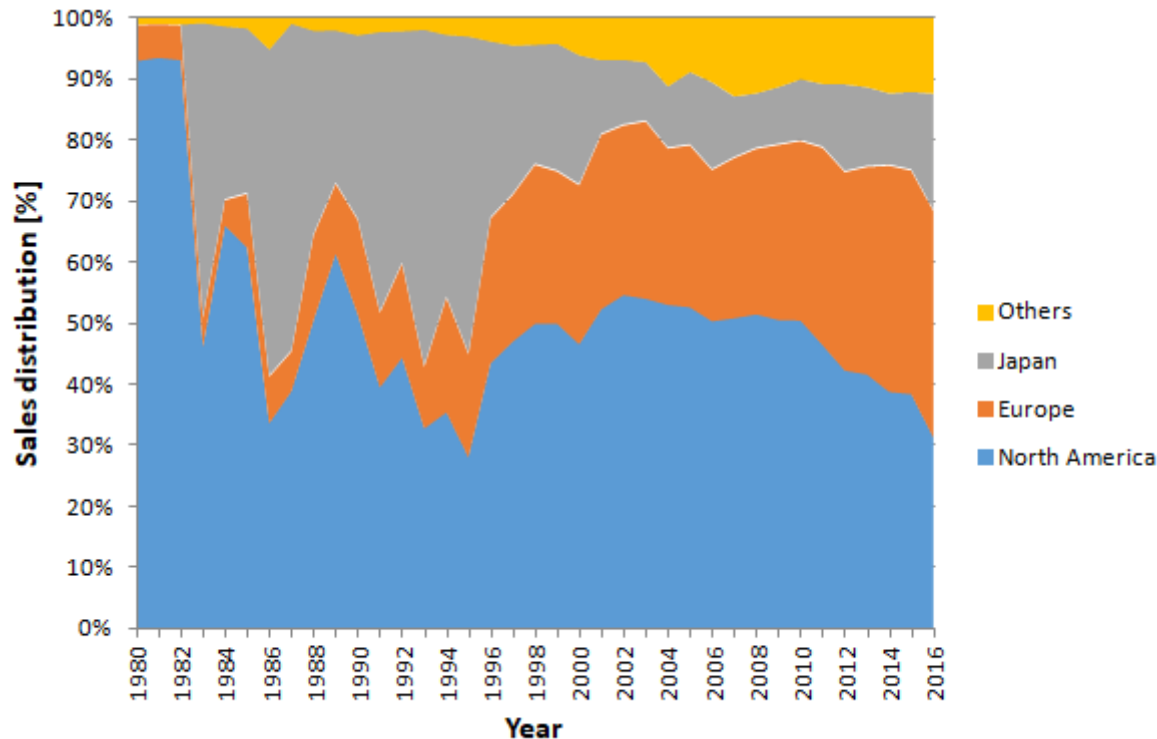


- The sales in each region changes over time, which challenges the present assumption that the sales keep constant over time.

Key visualization – 2



- Sales distribution for different regions



- North America: the sales in most cases are dominant large over all the time
- Europe: the sale's proportion keeps constantly around 1/3 in the last 10 years
- Japan: the sale's proportion generally keeps around 12%, but in 2016 is around 19%.
- Other regions: the proportion generally keeps around 13% in the last 10 years

Project: Preparing for influenza season

- Objective

Determine when to send staff, and how many to each state in USA

- Data sets:

‘influenza deaths by geography’ and
‘Population data by geography’

- Tools:

Excel and Tableau

The project report link and video link:

<https://public.tableau.com/app/profile/winter4019/viz/Projectpresentation-2-10/story-myproject>

<https://youtu.be/FMSQtpNndq8>, <https://youtu.be/WzAXaGDAxKc>

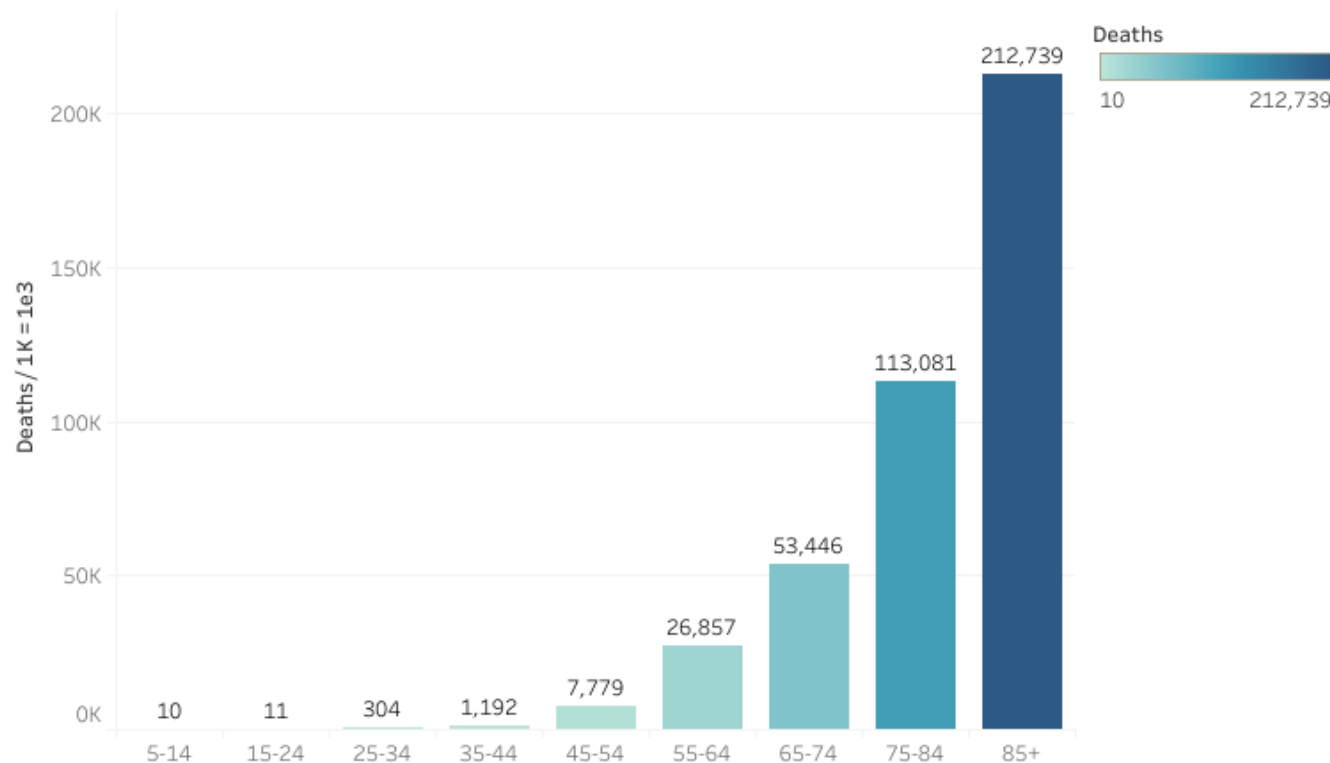
Analysis stages

Stages	Contents	Purpose
Verify the quality and integrity of data	Collect data, clean data, verify the completeness, accuracy and validity of data	To ensure the accuracy of the following analysis
Data analysis	Influenza deaths distribution in regions; Influenza deaths of vulnerable populations; influenza season length in different regions	To understand the features of influenza season, and its effects on the people
Modeling	Regression, Clustering and Forecasting	To have a better understanding and make predictions
Conclusion and recommendations	Get conclusion and give recommendations	To answer the key questions the stakeholders care

Key visualizations - 1

- Influenza deaths vs. age

Influenza deaths for different ten-year age groups (2009-2017)



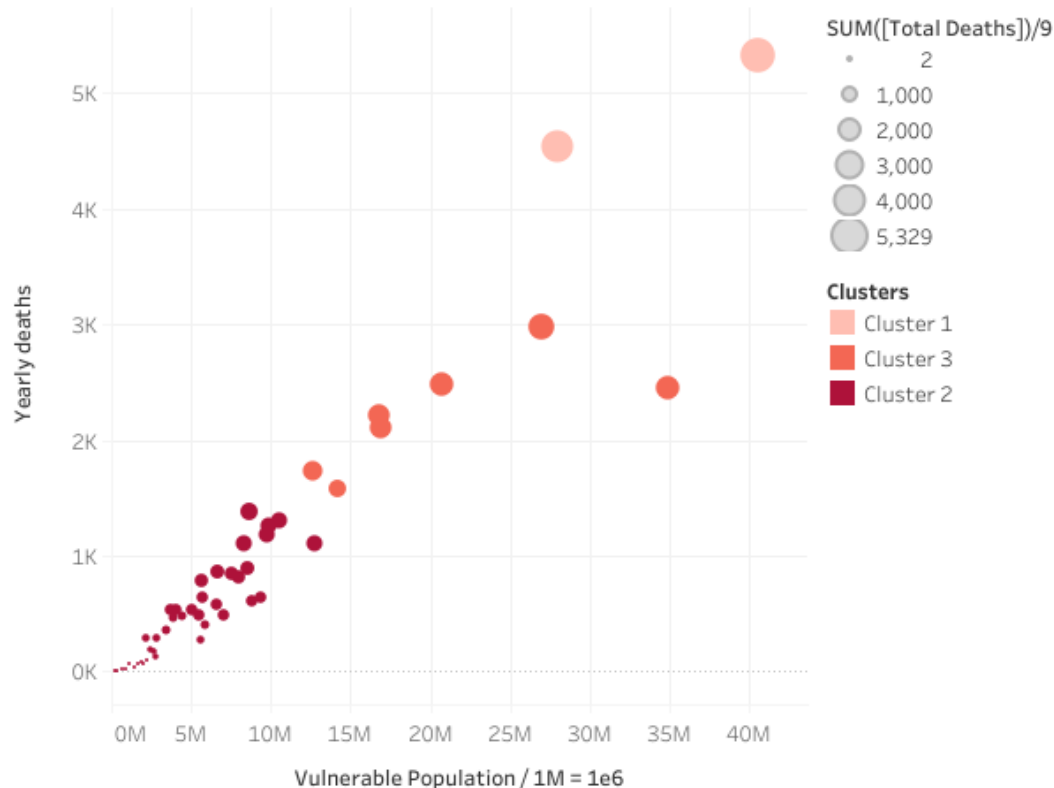
- Influenza deaths increases as the age increases; deaths of vulnerable population (age ≥ 65) makes up the biggest proportion of the total influenza deaths.

Key visualization – 2

- Population Clustering



States clusters based on scale of vulnerable population and yearly deaths of each state



Cluster 1 – two states that have the most yearly deaths and a large scale of vulnerable populations.

Cluster 3 – seven states that have a middle scale of yearly deaths and a middle scale of vulnerable populations.

Cluster 2 – other states that have a relatively small scale for both yearly deaths and vulnerable population.

Project: Rockbuster Stealth Video Rental

- Objectives:
use data to help with the launch strategy for the new online video service.
- Data sets:
'Rockbuster_data_set' that includes 15 different tables.
- Tools:
SQL and Tableau

The project report link:

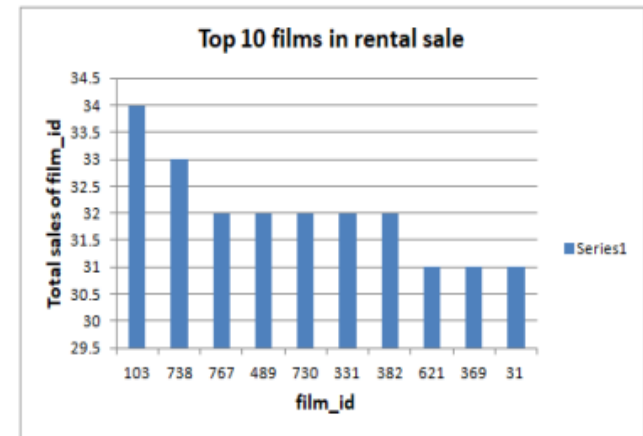
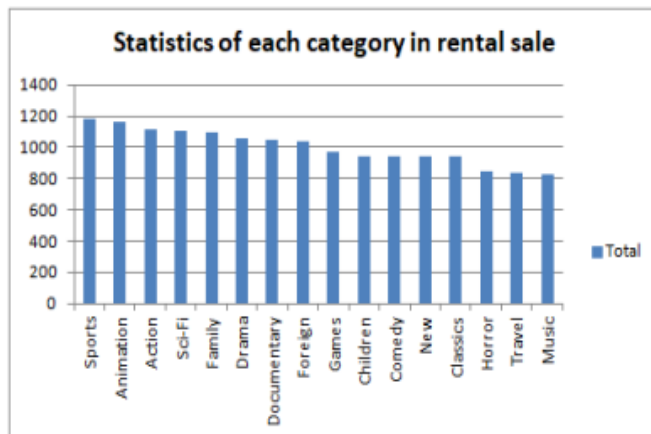
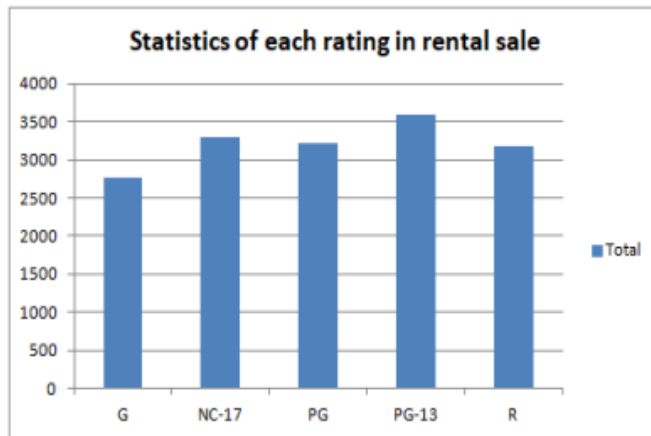
<https://public.tableau.com/app/profile/winter4019/viz/ProjectRockbuster/Story1>

Analysis stages

Stages	Contents	Purpose
Verify the quality and integrity of data	Collect data, clean data, verify the completeness, accuracy and validity of data	To ensure the accuracy of the following analysis
Data analysis	Descriptive statistics; Features of rating and category in rental sale; Customer distribution	To understand the features of rentals and differences in customers behaviors
Modeling	Clustering	To have a better understanding and make predictions
Conclusion and recommendations	Get conclusion and give recommendations	To answer the key questions the stakeholders care

Key visualizations - 1

- Statistics in rental sales



From these three Charts, one can conclude that

1) In the rental records, PG-13 is most popularly rented;

2) in the rental records, the top 5 film categories are Sports, Animation, Action, Sci-Fi and Family.

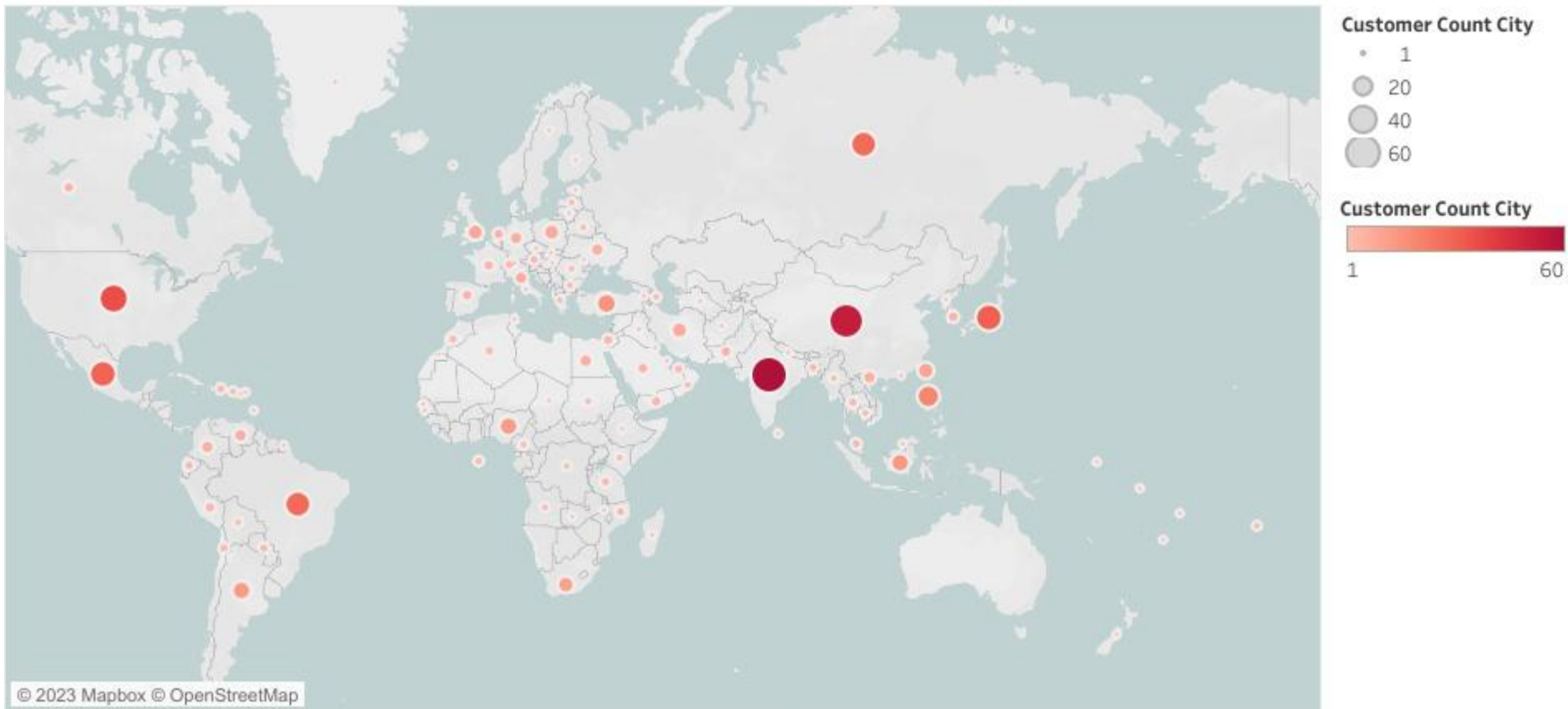
3) in the rental records, the top 10 film_ids are "Bucket Brotherhood", "Rocketeer Mother", "Juggler Hardly", "Grit Clockwork", "Ridgemont Submarine", "Forward Temple", "Scalawag Duck", "Zorro Ark", "Wife Turn" and "Hobbit ..

Key visualizations - 2



- Customer distribution all over the world

Customer counts and total payments in each country



- The top ten countries that have most customers are India, China, United States, Japan, Mexico, Brazil, Russian Federation, Philippines and Turkey.

Project: Instacart Grocery

- Objectives:

To uncover more information about the sales patterns

- Data sets

‘Customer_Data_Set’, ‘Data Dictionary’

- Tools:

Python, Excel

The project report link:

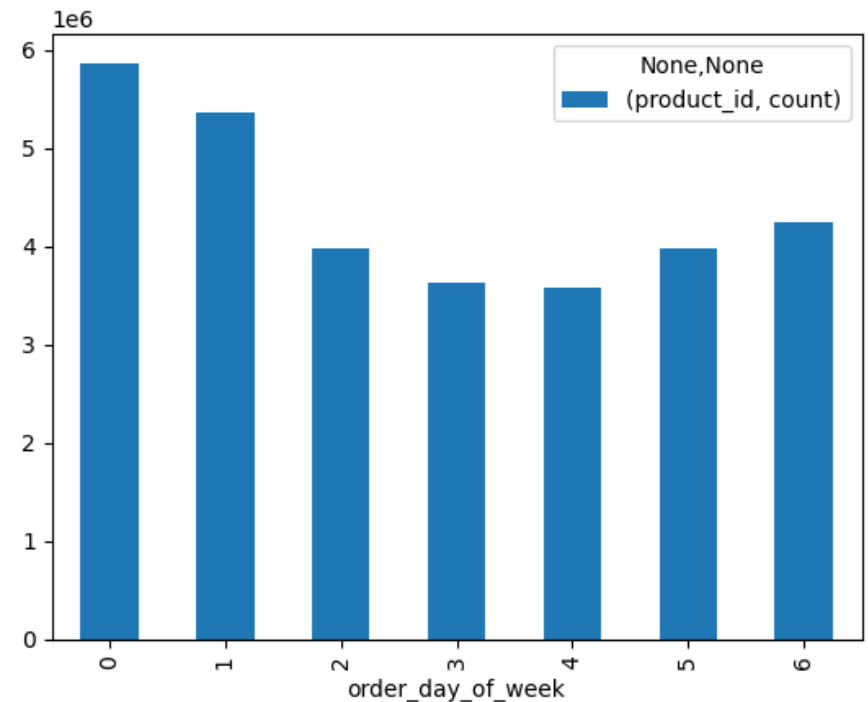
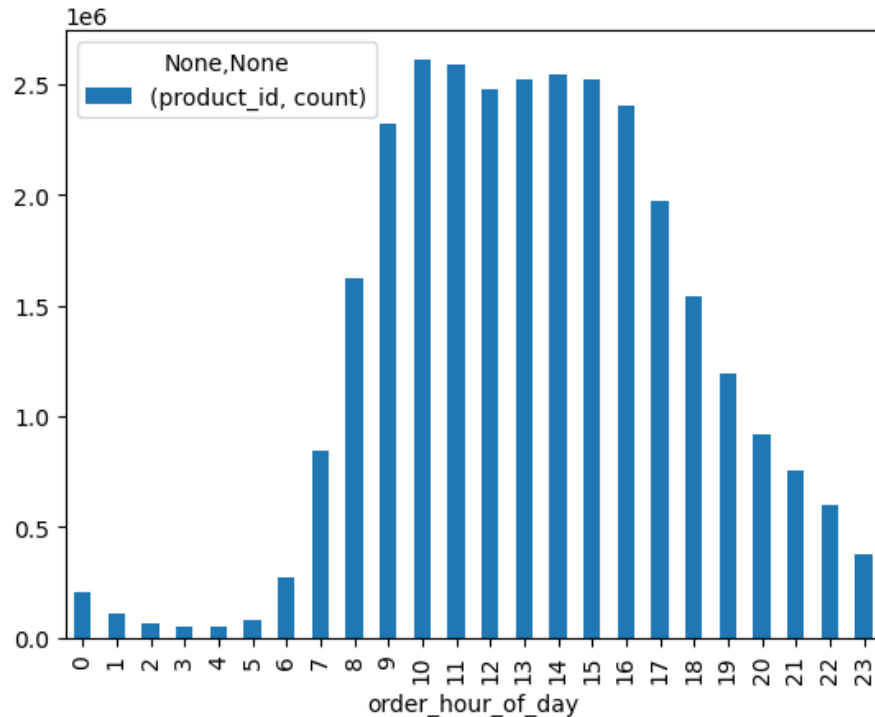
<https://github.com/dongmei-s/Project-Instacart/blob/main/4.10%20Codes%20in%20final%20report.ipynb>

Analysis stages

Stages	Contents	Purpose
Verify the quality and integrity of data	Collect data, clean data, verify the completeness, accuracy and validity of data	To ensure the accuracy of the following analysis
Data analysis	Customer consuming patterns over time (day and week); Customer consuming patterns over customer profiles;	To understand the customers better so that give target recommendations at different target time
Modeling	Regression, Clustering	To have a better understanding and make predictions
Conclusion and recommendations	Get conclusion and give recommendations	To answer the key questions the stakeholders care

Key visualizations - 1

- Customer consuming patterns over time

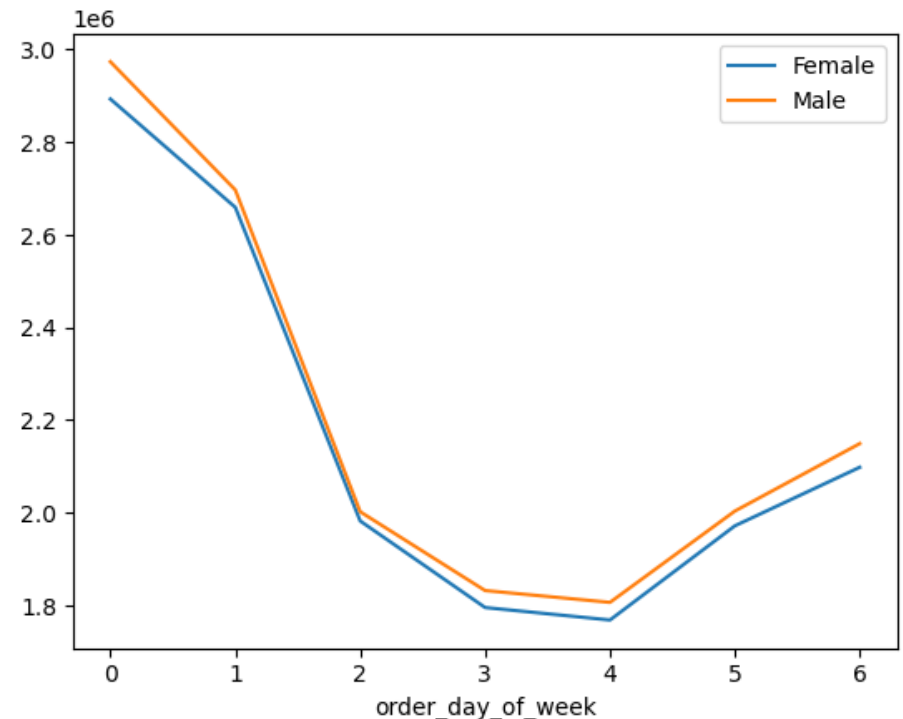
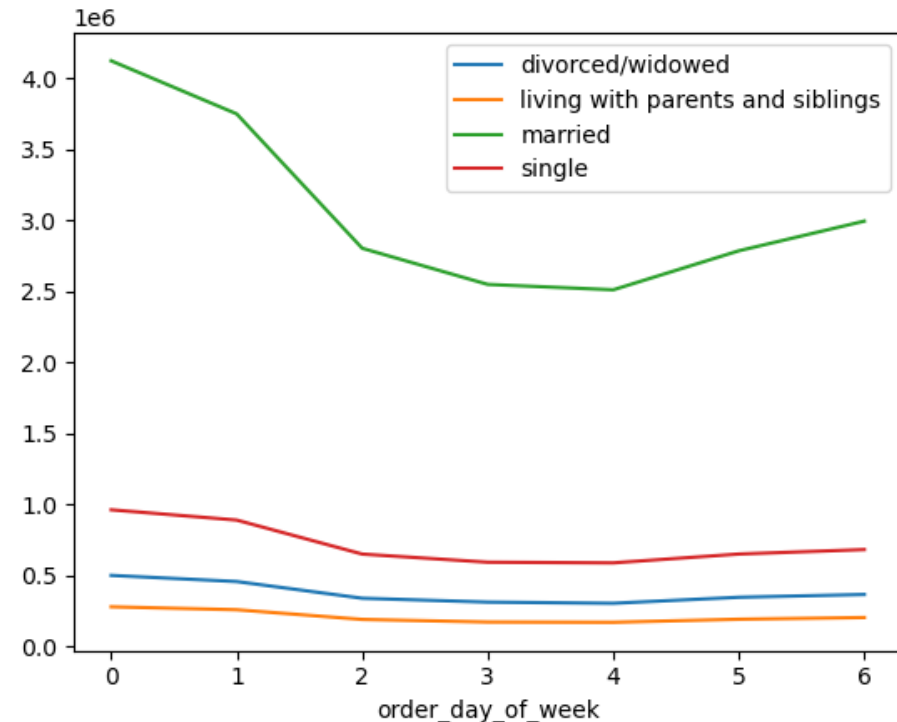


- Left figure: There are most consuming activities at around from 10 am to 16 pm, and least from 0 to 5 am.
- Right figure: There are most consuming activities on the day '0' and '1', and least on '3' and '4'.

Key visualizations – 2



- Customer patterns vs. Customer profiles



➤ Left figure: For the group 'married', the amount of products ordered by the customers changes over time most obviously, whereas for the group of 'living with parents and siblings', it doesn't change over time so obviously.

➤ Right figure: 1) The consuming patterns over day for female and male customers are consistent;

2) The male customers generally consume more than female customers.