Final Project - Zomato

(Schema, Pipeline Diagram, Tasks)

Team Members

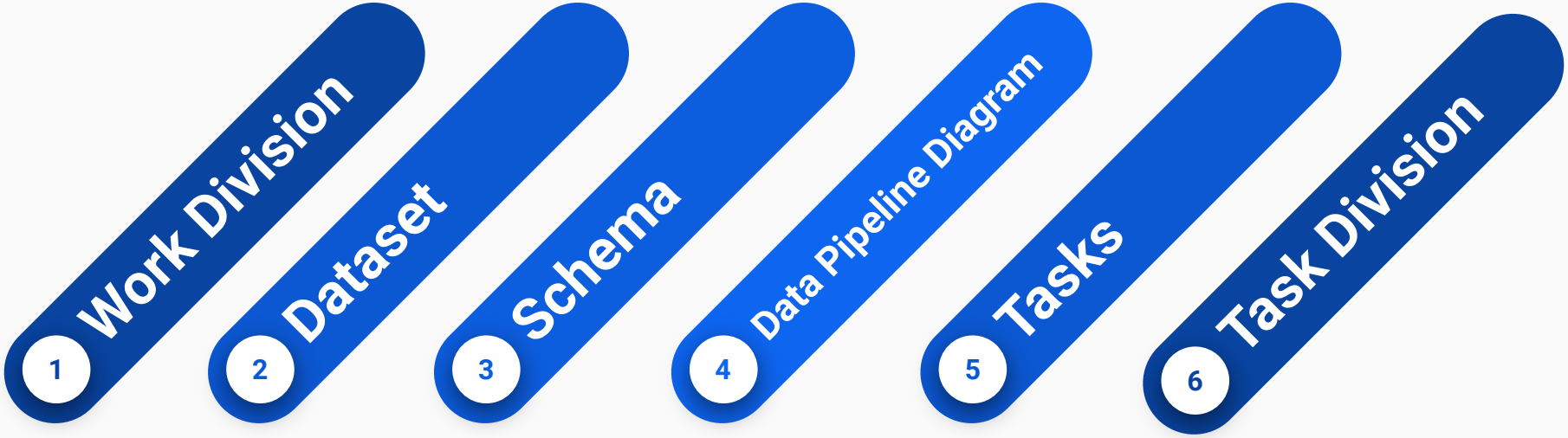
Baburam Shrestha

Pallavi Shrestha

Pradip Sapkota

Date: 2022-11-02

Contents


1
Work Division
2
Dataset
3
Schema
4
Data Pipeline Diagram
5
Tasks
6
Task Division

1. Work Division

Work	Assigned to
Searching and selecting dataset	All team members
Designing the schema of dataset	Pallavi Shrestha
Preparing data pipeline diagram	Pradip Sapkota
Preparing list of tasks to be done	Baburam Shrestha, Pradip Sapkota
Preparing presentation slides	Pallavi Shrestha
Tasks	5-6 tasks each
Integration	All team members

2. Dataset

- Name: Zomato
- URL: https://www.kaggle.com/datasets/rishikeshkonapure/zomato
- Size: 574.1 MB
- Rows: 51717
- Columns: 17
- Description: This dataset contains the data about the restaurants present in Zomato app database (an Indian restaurant aggregator and food delivery application).

Dataset (..contd)

zomato.csv (574.07 MB)

url	address	name	online_order	book_table	rate	votes	phone
https://www.zomato.com/bangalore/jalsa-banashankari?context=eyJzZSI6eyJlIjpbNTg2OTQsIjE4Mzc1NDc0IiwiwI...	942, 21st Main Road, 2nd Stage, Banashankari, Bangalore	Jalsa	Yes	Yes	4.1/5	775	080 42297555 +91 9743772233

location	rest_type	dish_liked	cuisines	approx_cost(f...	reviews_list	menu_item	listed_in(type)	listed_in(city)
Banashankari	Casual Dining	Pasta, Lunch Buffet, Masala Papad, Paneer Lajawab, Tomato Shorba, Dum Biryani, Sweet Corn Soup	North Indian, Mughlai, Chinese	800	[('Rated 4.0', 'RATED\n A beautiful place to dine in.The interiors take you back to the Mughal era....	[]	Buffet	Banashankari

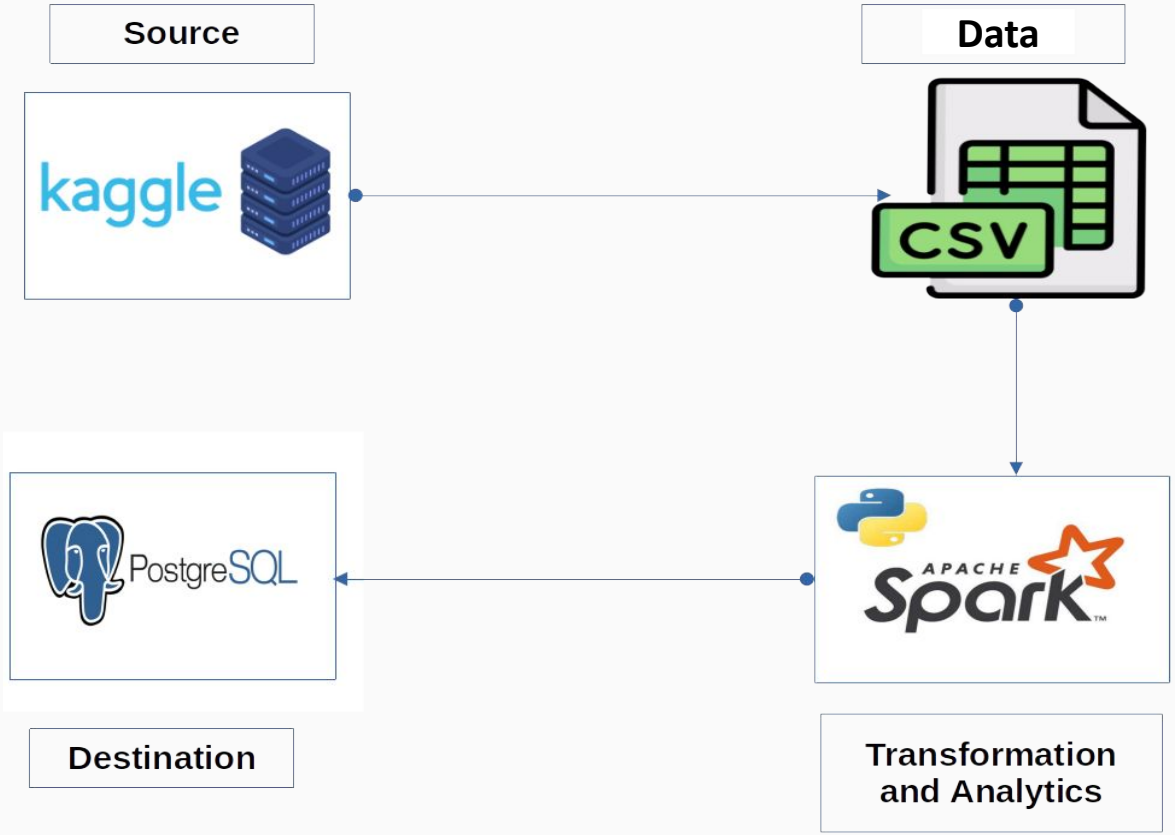
3. Schema

S.N.	Column	DataType	Description
1.	**url**	string	URL of the restaurant in zomato
2.	**address**	string	Complete address of the restaurant
3.	**name**	string	Name of the restaurant
4.	**online_order**	boolean	Do they accept online orders**?** (Yes/No)
5.	**book_table**	boolean	Can we book a table at the restaurant? (Yes/No)
6.	**rate**	string	Rating given to the restaurant on the zomato app
7.	**votes**	integer	Number of people who gave the rating
8.	**phone**	string	Phone Number of the restaurant
9.	**location**	string	Area of the restaurant (area in the address)

Schema (..contd)

S.N.	Column	DataType	Description
10.	**rest_type**	string	Type of the restaurant (Casual Dining/Cafe/Quick….)
11.	**dish_liked**	string	Dishes liked by most people
12.	**cuisines**	string	Cuisines served by the restaurant
13.	**approx_cost (for two people)**	float	Approximate Cost for Two people
14.	**reviews_list**	string in JSON	List of reviews received by the restaurant (JSON)
15.	**menu_item**	string in List	List of items in the restaurant’s menu
16.	**listed_in(type)**	string	Category of restaurant(Delivery/ Dine-Out/Buffet etc...)
17.	**listed_in(city)**	string	Name of the city

4. Data Pipeline Diagram



5. Tasks

1. Find the details of the restaurant that has the facility of a “book table” before.
2. Suggest where one can open a new restaurant (location).
3. Compare the restaurant to whether it has the facility of an “online order” or not.
4. Make a group of restaurants according to the type of restaurant and location of the restaurants.
5. List the top ten restaurants in a given city.
6. List the top ten restaurants with the highest number of branches.
7. Count the number of restaurants that allows online orders
8. List the top ten restaurants with the highest rating
9. List restaurants that are either cafes or Quick Bites
10. List the best restaurants in each location(best in accordance with rate)
11. Find the average, max, and min of votes grouped by location
12. Change yes to true and Not to False in columns online_order and book_table using UDP.
13. Most liked dishes in the city Banashankari
14. Change rate to float type by removing ‘/5’
15. Find the total no. of unique dishes found in each city (window function)

6. Task Division

1. **Baburam Shrestha**
    a. Find the restaurants which serve Italian items.
    b. Most liked restaurant in the city Banashankari
    c. Change yes to True and Not to False in columns online_order and book_table using UDF.
    d. List the best restaurants in each location(best in accordance with rate
    e. Find the total no. of voters in each cities (window function)

Task Division (..contd)

2. **Pallavi Shrestha**
   a. Find the details of the restaurant that has the facility of a “book table” before.
   b. Make a group of restaurants according to the type of restaurant and location of the restaurants.
   c. Change rate to float type by removing ‘/5’
   d. List the top ten restaurants with the highest rating
   e. Find the average, max, and min of votes grouped by location
   f. Find the correlation between approx_cost_two_people and rating of the restaurants.

Task Division (..contd)

3. Pradip Sapkota

a. Suggest where one can open a new restaurant

b. Compare the restaurant to whether it has the facility of an “online order” or not.

c. List the top ten restaurants with the highest number of branches.

d. List restaurants that are either cafes or Quick Bites

e. Count the number of restaurants that allows online orders and book table.

f. Get the number of restaurants in each city.

THANK YOU!