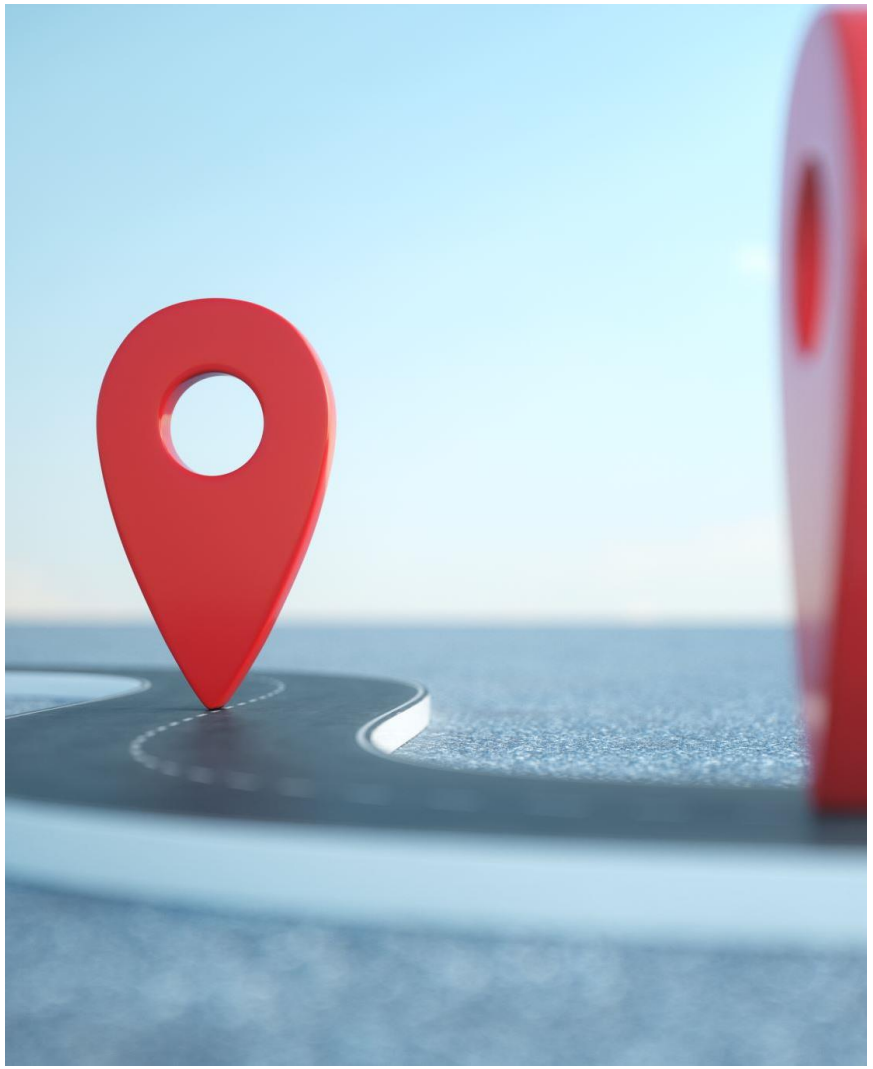


LOCATION BASED RESTAURANTS RECOMMENDATION SYSTEM

Data 603 – Platforms for Big Data Processing

Prof. Dr. Najam Hassan

Group Members:

- Adarsh Rao Akula
- Chandana Anugu
- Satvik Reddy Singi Reddy

INTRODUCTION

In today's digitally-driven world, personalized recommendations are crucial for businesses to thrive. By leveraging big data analytics, we aim to revolutionize restaurant recommendations, providing users with tailored suggestions based on their preferences.

Significance of the Project:

- Utilizes advanced data tools to analyze Yelp reviews and locations.
- Focuses on providing restaurant recommendations based on geographic insights.

Relevance to the Course:

- Applies classroom theory to real-world data problems.
- Covers the entire process: data ingestion, processing, analysis, and visualization.

WHY DO WE NEED A RECOMMENDATION SYSTEM?

For Users:

Speed up searches

Easy to access the content they might be interested in

Receive many offers they would have never searched for

For Restaurants:

Attracts customers

Competitive advantage by reducing the threat of losing their customers to their competitors

Increase their earnings

PROJECT OBJECTIVES:

- Our approach involves developing a personalized recommendation system that suggests restaurants based on user preferences.
- To achieve this, we'll leverage powerful big data platforms such as Apache Spark and PySpark.
- By efficiently processing and analyzing the large-scale datasets from Yelp, we aim to provide users with tailored suggestions that enhance their overall experience and satisfaction with the platform.

Specific Problems Addressed:

- Leveraging Apache Spark and PySpark to handle big data processing tasks efficiently ensures scalability and high performance, which is crucial given the volume and complexity of datasets from platforms like Yelp.
- Implementing data preprocessing techniques is crucial for ensuring the quality and consistency of data, which in turn is essential for accurate analysis and reliable recommendations.

DATA SOURCES AND COLLECTION:

Data Sources Used:

The Yelp dataset, with its varied data types like business details, user reviews, and profiles, served as an ideal source for comprehensive analysis.

Data Collection Process:

1. Ensured compliance with data usage policies and regulations when accessing the Yelp dataset.
2. Data Extraction.
3. Data Storage.

Challenges Encountered:

1. Encountered challenges related to the availability and accessibility of specific data attributes within the Yelp dataset.
2. Data Format Inconsistencies.
3. Dealt with the sheer volume of data within the Yelp dataset, requiring efficient storage and processing solutions to handle large-scale data operations.

DATASET INFORMATION

- Business dataset: 111.9MB
- Review dataset: 5.53GB
- User dataset: 3.36GB

```
Business Count: 150346
Reviews Count: 6990280
Users Count: 1987897
+--------------------+--------------------+--------------------+--------------------+----------+------
----+-------+----------+-----------+--------------------+----------+-----------+-----+-----+
|             address|          attributes|         business_id|          categories|      city|
ours|is_open|  latitude|   longitude|                name|postal_code|review_count|stars|state|
+--------------------+--------------------+--------------------+--------------------+----------+------
----+-------+----------+-----------+--------------------+----------+-----------+-----+-----+
|1616 Chapala St, ...|{NULL, NULL, NULL...|Pns2l4eNsfO8kk83d...|Doctors, Traditio...|Santa Barbara|
NULL|      0|34.4266787|-119.7111968|Abby Rappoport, L...|      93101|          7|  5.0|   CA|
|87 Grasso Plaza S...|{NULL, NULL, NULL...|mpf3x-BjTdTEA3yCZ...|Shipping Centers,...|       Affton|{8:0-1
0...|      1| 38.551126|  -90.335695|       The UPS Store|      63123|         15|  3.0|   MO|
|5255 E Broadway Blvd|{NULL, NULL, NULL...|tUFrWirKiKi_TAnsV...|Department Stores...|       Tucson|{8:0-2
2...|      0| 32.223236| -110.880452|              Target|      85711|         22|  3.5|   AZ|
|         935 Race St|{NULL, NULL, u'no...|MTSW4McQd7CbVtyjq...|Restaurants, Food...| Philadelphia|{7:0-2
0...|      1|39.9555052| -75.1555641|  St Honore Pastries|      19107|         80|  4.0|   PA|
|       101 Walnut St|{NULL, NULL, NULL...|mWMc6_wTdE0EUBKIG...|Brewpubs, Breweri...|   Green Lane|{12:0-
L,...|      1|40.3381827| -75.4716585|Perkiomen Valley ...|      18054|         13|  4.5|   PA|
+--------------------+--------------------+--------------------+--------------------+----------+------
----+-------+----------+-----------+--------------------+----------+-----------+-----+-----+
only showing top 5 rows

+--------------------+----+--------------------+-----+--------------------+-----+--------------------+----
------------+
|         business_id|cool|                date|funny|           review_id|stars|                text|usef
user_id|
+--------------------+----+--------------------+-----+--------------------+-----+--------------------+----
------------+
|XQfwVwDr-v0ZS3_Cb...|   0|2018-07-07 22:09:11|    0|KU_O5udG6zpxOg-Vc...|  3.0|If you decide to ...|
6K5RLWhZyI...|
|7ATYjTIgM3jUlt4UM...|   1|2012-01-03 15:28:18|    0|BiTunyQ73aT9WBnpR...|  5.0|I've taken a lot ...|
OKpv6SyGZT...|
|YjUWPpI6HXG530lwP...|   0|2014-02-05 20:30:30|    0|saUsX_uimxRlCVr67...|  3.0|Family diner. Had...|
SiwikVnbP2...|
|kxX2SOes4o-D3ZQBk...|   1|2015-01-04 00:01:03|    0|AqPFMleE6RsU23_au...|  5.0|Wow!  Yummy, diff...|
Uuf5__HHc_...|
|e4Vwtrqf-wpJfwesg...|   1|2017-01-14 20:54:15|    0|Sx8TMOWLNuJBWer-0...|  4.0|Cute interior and...|
dDog4jkNY9...|
+--------------------+----+--------------------+-----+--------------------+-----+--------------------+----
------------+
only showing top 5 rows
```

TOOLS AND TECHNOLOGIES

Big Data Platforms and Frameworks:

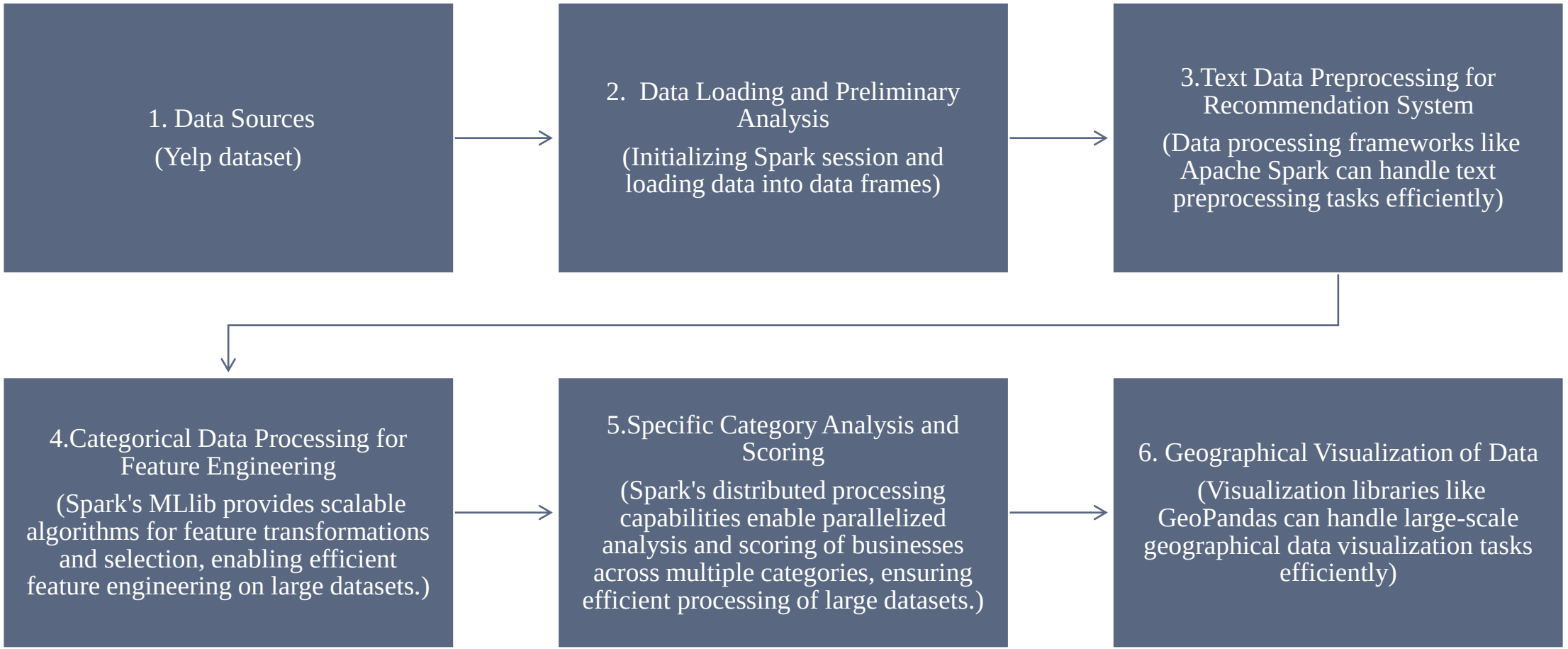
1. **Apache Spark:**
   - Leveraged Apache Spark for its scalability, performance, and support for complex data processing tasks.
   - Utilized Spark's RDDs (Resilient Distributed Datasets) and DataFrame APIs for distributed data processing.
   - Apache Spark was chosen for its ability to handle large-scale datasets efficiently, ensuring scalability and high performance.
2. **MLlib:**
   - Integrated MLlib for developing machine learning models and pipelines, enabling tasks such as text analysis and recommendation system building.
   - MLlib was preferred for its extensive set of machine learning algorithms and distributed computing capabilities, enabling the development of advanced recommendation systems.
3. **PySpark:**
   - Employed PySpark to leverage Spark's functionalities within the Python ecosystem.

CONTENT-BASED FILTERING

- Content-based filtering is useful in text analysis tasks like generating restaurant recommendations based on review content, where it can quantify the similarity between text data, allowing the system to identify and recommend restaurants with similar features or reviews. Uses distance metrics to evaluate how similar 2 items are based on different users' feedback.

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

STACK DIAGRAM



REPORTS, INSIGHTS, AND RECOMMENDATIONS:

Reports:

- Analyzed user interactions with restaurants, including ratings, reviews, and preferences, and evaluated the performance of restaurants based on those factors.
- Identified top-performing restaurants and understood market trends and consumer preferences.

Key Insights:

- Identified specific cuisines, ambiance preferences, and service attributes favored by users.
- Uncovered correlations between restaurant attributes (e.g., cuisine type, location) and user ratings, providing valuable insights.

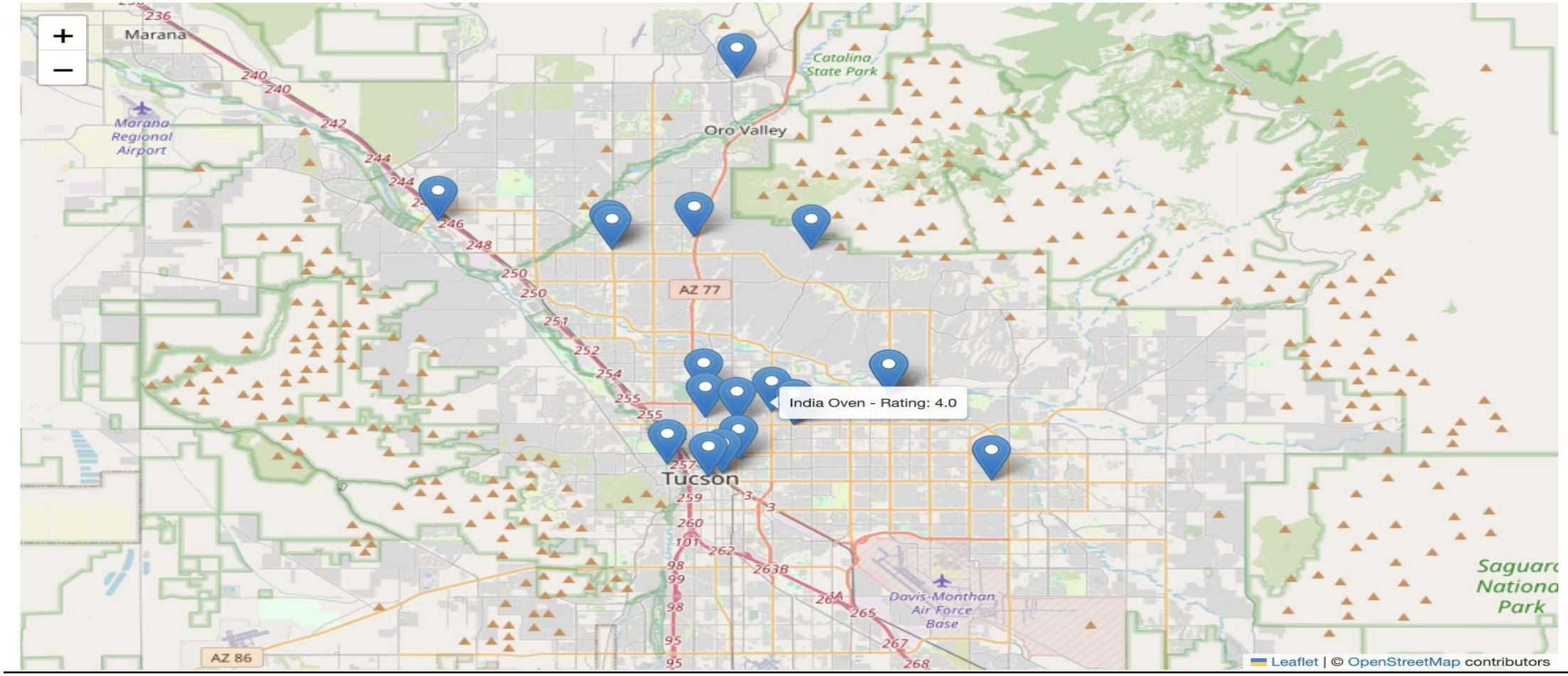
Recommendations for C-Level Executives:

- Invest in personalized recommendation systems leveraging machine learning algorithms to enhance user engagement and satisfaction.
- Forge Strategic Partnerships.

RESTAURANT RECOMMENDATION:

```
Restaurants similar to keyword - Indian
+----------------------+----------------+-----------------------------------+-----------------------------------
-----------------------------------------------------------------------------------------------------+-----+------
------+
|business_id           |similarity_score|name                               |categories
|stars|review_count|
+----------------------+----------------+-----------------------------------+-----------------------------------
-----------------------------------------------------------------------------------------------------+-----+------
------+
|TeDC6CPQS2AsgEk7TFi5ag|0.9             |Indian Queen Lounge & Bar          |Indian, Restaurants, Hookah Bars, Ba
rs, Nightlife, Lounges, Buffets, Chicken Wings                                                       |3.5  |11
|
|J_3vgev5HjjFSHDinRrA-A|0.9             |Masala Wok                         |Restaurants, Indian, Chinese, Vegeta
rian                                                                                                 |2.0  |33
|
|IqitQsWPnTVTCJIRqh80lA|0.9             |India Bistro                       |Restaurants, Indian
|4.0  |186         |
|XEsS_1FRtikXa2WPs3gzkQ|0.9             |Paradise Biryani Pointe            |Halal, Food, Restaurants, Indian
|4.0  |74          |
|-0TffRSXXIlBYVbb5AwfTg|0.9             |IndeBlue Modern Indian Food & Spirits|Cocktail Bars, Food Delivery Service
s, Nightlife, Breakfast & Brunch, Food, Bars, Event Planning & Services, Caterers, Restaurants, Indian|4.5  |1097
|
|4r4eZQBqks0Ky_OHK-LU7Q|0.9             |Bombay Grill                       |Indian, Buffets, Pakistani, Restaura
nts                                                                                                  |4.0  |192
|
|9dL1rsPANYr-71hdwoY-CA|0.9             |Desi Tadka Indian Cuisine          |Restaurants, Indian
|4.0  |169         |
|QhJCaPpJT1iMx1X9JGxJJg|0.9             |The Tandoor                        |Restaurants, Imported Food, Indian,
Ethnic Food, Specialty Food, Food                                                                    |3.5  |15
|
|f82dhKNiUXsDVPMLqKYiIQ|0.9             |Sher-e-Punjab                      |Restaurants, Salad, Pakistani, India
n, Cocktail Bars, Food, Food Delivery Services, Soup, Halal, Bars, Nightlife                         |4.0  |446
|
|Q-fRS0NkDmFTCBUNbpERxg|0.9             |Chapati Beta                       |Restaurants, Indian, Pakistani, Hala
l                                                                                                    |4.5  |15
|
+----------------------+----------------+-----------------------------------+-----------------------------------
-----------------------------------------------------------------------------------------------------+-----+------
```

VISUALIZATION:



FUTURE OPPORTUNITIES AND CHALLENGES:

Opportunities:

1. Enhanced Personalization: Integrate additional user data sources like social media activity, location data, and demographic information to bolster personalization capabilities, ensuring tailored experiences for users.

2. Diversification of Revenue Streams: Offer additional services such as event planning, food delivery, or culinary experiences based on customer preferences to diversify revenue streams and expand business offerings.

Challenges:

1. Privacy and Ethical Considerations: Address potential privacy and ethical concerns associated with the utilization of advanced analytics and AI techniques on user data to maintain user trust and compliance with regulations.

2. Technical Challenges: Overcome technical hurdles related to real-time data processing and system latency to deliver seamless user experiences, ensuring optimal performance and responsiveness of the platform.

CONCLUSION

1.**Effective Use of Technology**: The project successfully used tools like Apache Spark to analyze large amounts of data from Yelp. This helped create a system that recommends restaurants based on what users like and where they are located, showing how classroom learning can be applied to solve real problems.

2. **Insights for Businesses:** By examining how users interact with restaurants, the project uncovered trends and preferences that can help businesses improve how they engage with customers. This could give them an advantage by making recommendations more personal and relevant.

3. **Looking Ahead**: The project pointed out opportunities to make the recommendation system even better by using more data sources and offering new services. It also noted the importance of handling privacy carefully and improving the system’s ability to process data quickly to keep users happy and trustful.

REFERENCES

- Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, R. (2020). Recommendation systems: algorithms, challenges, metrics, and business opportunities. *Applied Sciences*, *10*(21), 7748. https://doi.org/10.3390/app10217748

- Ko, H., Lee, S., Park, Y., & Choi, A. (2022). A survey of recommendation systems: recommendation models, techniques, and application fields. *Electronics*, *11*(1), 141. https://doi.org/10.3390/electronics11010141