Group D

Analytics on YELP

Data Technology Solutions

“In God we trust. All others must bring data.”

Professor W. Edwards Deming

Engineer, Statistician

Developed the sampling techniques

GRATITUDE

We like to thank our teacher Mr. Bhavik Gandhi for sharing his knowledge and ideas. We are fortunate to have him as our teacher for this course

VP, Product and Engineering, HaikuJAM
Data Science Manager, Shopify
Associate Vice President, Service & Analytics, Pepperfry
Program Manager and Software Developement Engineer, Microsoft

"Are marriages made in heaven or do algos decide who we love?"
Talk about 3 different use cases of data science at Shaadi.com
https://www.youtube.com/watch?v=sJIfHYDFtMk

Midpoint
OBJECTIVES

Objective 1

The objective of this project is to help Yelp through analyzing the Yelp dataset and gain insights into various aspects of businesses, users, and reviews.

Objective 2

The analysis aims to extract valuable information such as top-rated businesses, user behaviors, popular business categories, geographic trends, and more.

Objective 3

The project aims to use data exploration, visualization, and statistical analysis techniques to uncover patterns and trends within the Yelp dataset.

WHAT IS YELP

Yelp is a user-generated review platform that operates in the local business review and recommendations industry. Its primary focus is on providing a platform for users to discover and review local businesses, while also offering businesses an opportunity to engage with their customers and potentially attract new ones.

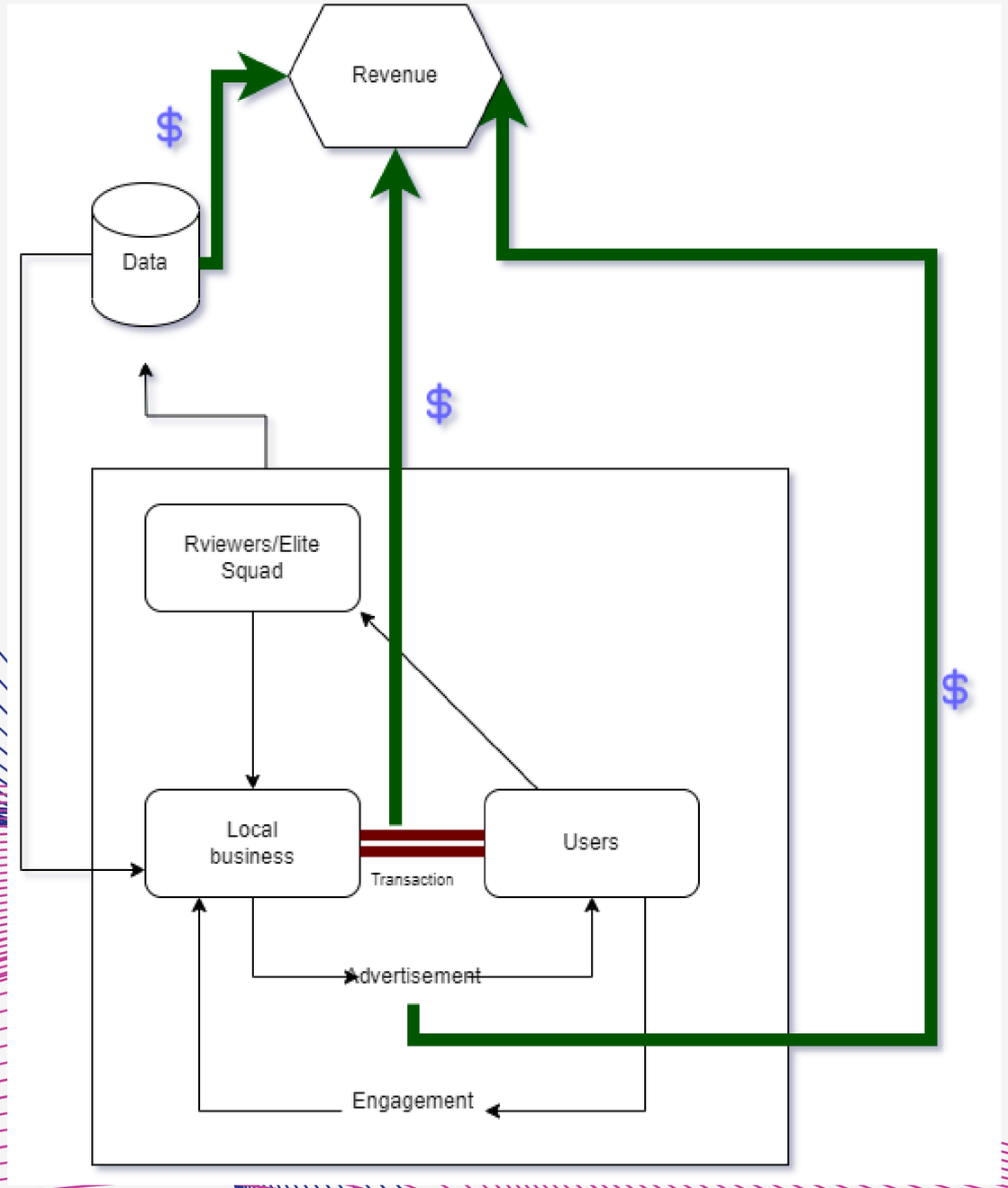
HOW YELP WORKS

- **User-Generated Content**: The core of Yelp's platform is driven by user-generated reviews and recommendations, providing valuable insights for other users to make informed decisions.
- **Advertising Revenue Model**: Yelp's significant revenue stream comes from businesses paying for advertising services, increasing their visibility on the platform.
- **Mobile App Optimization**: Yelp's mobile app plays a crucial role in its success, enabling users to access reviews and discover businesses while on the go, leveraging location-based services for real-time recommendations.
- **Community Engagement**: Yelp fosters an active community of reviewers, known as "Yelpers," contributing to the platform's credibility and attracting new users through genuine user experiences.

UNDERLYING REASONS FOR YELP SUPREMACY

:

- Large and Active Community: One of the primary reasons to choose Yelp is its large and active user community. With millions of users leaving genuine reviews and ratings, businesses can tap into a vast audience and increase their visibility.
- Authentic and Trustworthy Reviews: Yelp's user-generated reviews are authentic and unbiased, giving businesses and users confidence in the credibility of the feedback. This authenticity helps users make informed decisions about where to spend their time and money.
- All-in-One Business Information: Yelp provides comprehensive business information, including contact details, photos, menus, and operating hours. Having all this data in one place makes it convenient for users to find everything they need about a business quickly.
- User-Friendly Mobile App: Yelp's mobile app is easy to use and feature-rich, making it a preferred choice for users searching for local businesses on the go. The app's convenience ensures that users can access reliable information wherever they are.
- Refined Search Filters: Yelp's advanced search filters allow users to narrow down their search results based on specific criteria, such as location, category, price range, and ratings. This feature helps users find businesses that suit their preferences.



Business Process MAPPING

BUSINESS MODEL CANVAS yelp

Key Partners

- Content Contributors (Active reviewers)
- Partnerships
- Extended Partners

Key Activities

- Network Effects
- Customer Experience
- Ad Selling
- Platform Growth
- Grow local communities
- Raise brand awareness
- Manage day to day operation

Key Resources

- Employees
- High-Quality Content
- Brand Reputation
- Engaged Communities
- Partnerships

Value Propositions

local Businesses:

- exposure and visibility
- Increase traffic
- Reviews to promote businesses
- Targeted advertisements

Users:

- browse local businesses
- discover new places
- save time and money

Content Contributors:

- Engage in fun social interaction
- Support the community and local businesses

Customer Relationships

- Self Service Platform
- Brand Reputation
- Engagement
- Dedicated Sales
- Engaged Community
- Communication

Channels

- website or app
- Ads channel
- webpages for the business owner and developers
- Social media pages

Customer Segments

- Local Businesses
- Users
- Content contributors (Active reviewers)

Cost Structure

- Cost of Revenue (6%)
- Sales and Marketing (51%)
- Product Development (23%)
- General and Administrative (13%)
- Depreciation (5%)

Revenue Streams

- Advertisement (97% of revenue)
- Transaction revenue (1% of revenue)
- Other Services (2% of revenue)

Business Strategy Hub

source :bstrategyhub.com

Business REGISTRATION

Before registration process there are three requirements that are to be fulfilled:

- **Identify business structure**: There are several types of business structures like sole, partnership, each with their own costs, risks, and tax implications.
- **Select the business locations**: The registration process for businesses varies from state to state, and so do the business taxes must pay. As such, selecting business location is a significant decision to make.
- **Pick the business name**: A business name that want to stand out from the crowd. A creative, descriptive, or fun business name can make a lasting impression on potential customers.

How Yelp
MAKES BUSINESS FROM BUSINESSES

01

Businesses can offer deals and gift certificates through Yelp’s platform. These promotions attract customers and drive sales, with Yelp taking a percentage of the revenue generated from these offerings.

02

Yelp offers a reservation and waitlist management system forrestaurants. Participating restaurants pay a fee for using these services, which help streamline the reservation process for customers and generate revenue for Yelp.

05

Yelp offers a reservation and waitlist management system forrestaurants. Participating restaurants pay a fee for using these services, which help streamline the reservation process for customers and generate revenue for Yelp.

03

Businesses can pay for sponsored placements on Yelp's platform to increase visibility and reach a larger audience.

04

Businesses can use Yelp connect to share updates with their followers, and Yelp may charge for this feature.

Opportunities We Are Addressing

VALUE PROPOSITION

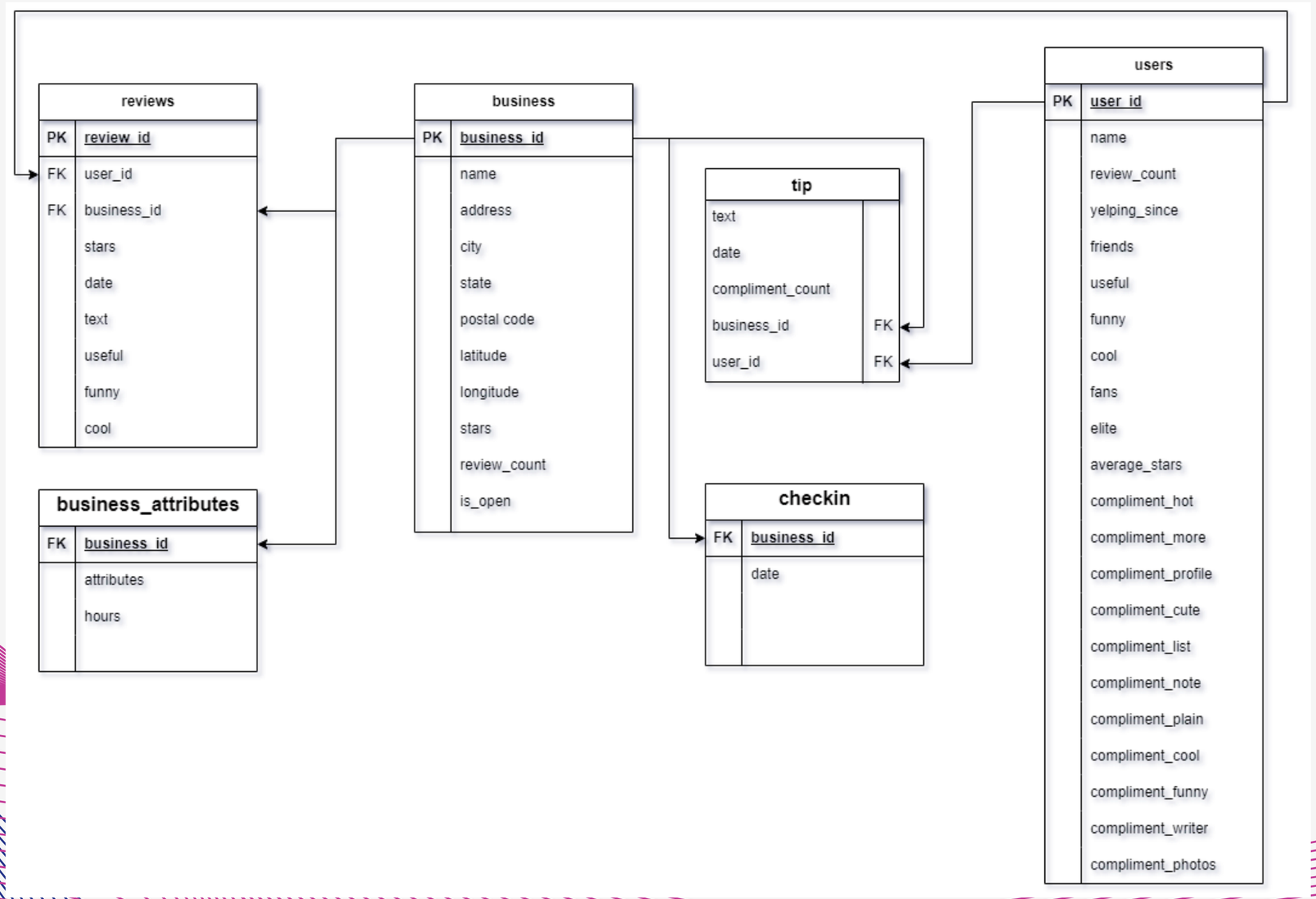
- **Business Optimization**: The actionable insights derived from our analytics project enable Yelp to identify inefficiencies and bottlenecks within their processes, allowing them to optimize operations and improve overall productivity.
- **Customer Behavior Analysis**: Through the analysis of customer data, we can provide actionable insights on consumer behavior, preferences, and pain points, enabling Yelp to tailor their products and services to better meet customer needs.
- **Product Innovation**: Our analytics project highlights areas for potential product improvements and innovation.
- **Risk Mitigation**: Actionable insights from our analytics efforts help the organization to proactively identify potential risks and vulnerabilities, allowing them to implement strategies to mitigate these risks and ensure long-term stability and success.

Data Model & Analytics

Data
SOURCE

- The data source for this project is the Yelp dataset, which is a collection of data related to businesses, users, and reviews from the Yelp platform.
- The dataset includes information about businesses such as their names, categories, locations, and star ratings. It also contains user profiles with details like user IDs, names, review counts, and average star ratings. Additionally, the dataset includes individual reviews that users have left for businesses, along with associated star ratings.
- Link : https://www.yelp.com/dataset
- Size : 9.29 GB

Data Model


reviews
PK review_id
FK user_id
FK business_id
stars
date
text
useful
funny
cool
business
PK business_id
name
address
city
state
postal code
latitude
longitude
stars
review_count
is_open
users
PK user_id
name
review_count
yelping_since
friends
useful
funny
cool
fans
elite
average_stars
compliment_hot
compliment_more
compliment_profile
compliment_cute
compliment_list
compliment_note
compliment_plain
compliment_cool
compliment_funny
compliment_writer
compliment_photos
tip
text
date
compliment_count
business_id FK
user_id FK
business_attributes
FK business_id
attributes
hours
checkin
FK business_id
date

PROJECT CHALLENGES

LARGE DATASET

The Yelp dataset is quite large, and handling big data efficiently is crucial. We tackled this challenge by using pandas for reading and cleaning the data and SQLite database for storing the data and run SQL queries.

DATA CLEANING

Datasets may contain missing or inconsistent data, requiring careful data cleaning and preprocessing. We used pandas to handle missing values and ensure data quality.

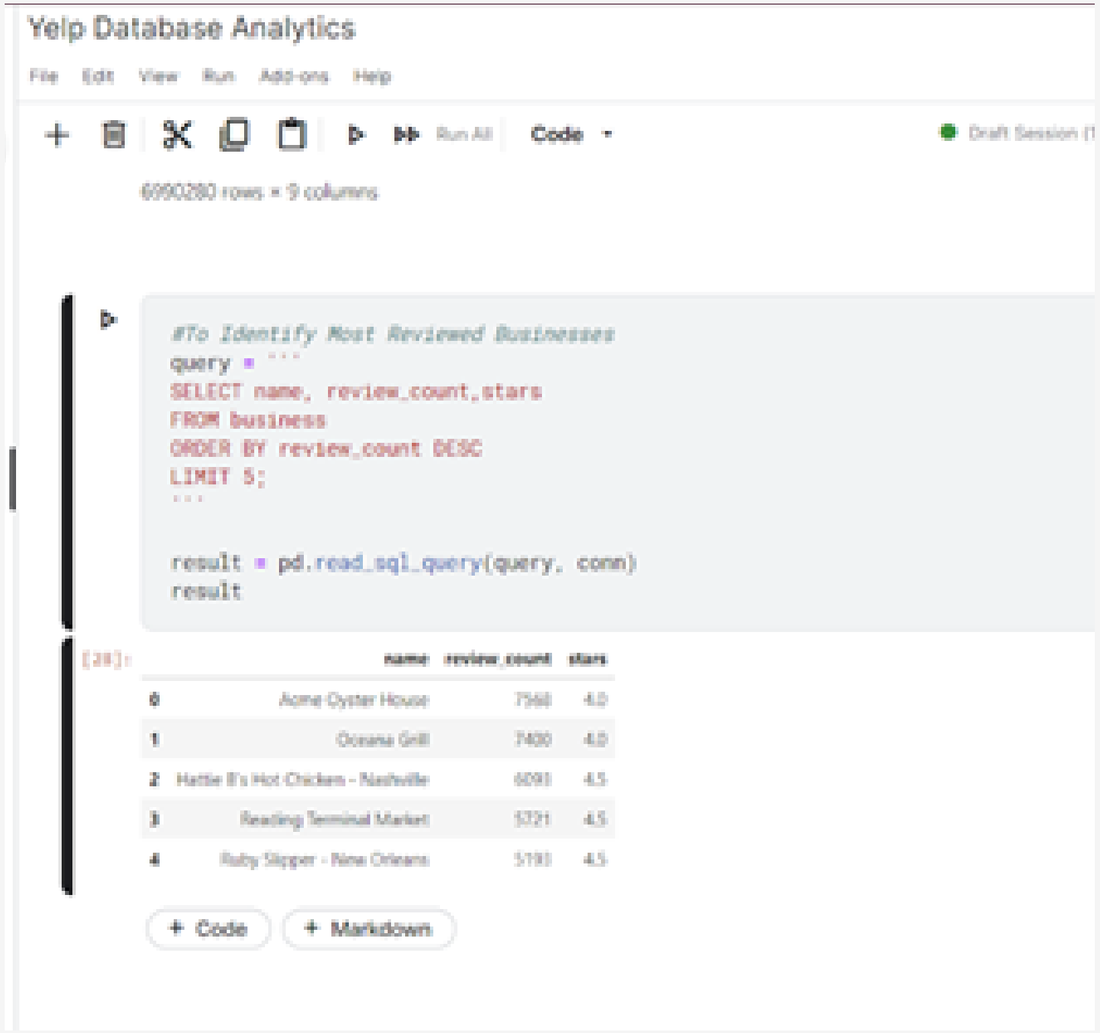
JSON FILES

All of the original files were in JSON formats and those were not parsed by Postgres and Bigquery. We moved to SQLite through Python to push the data to the database and running queries

PROCESSING AND STORAGE

Due to our processing power limitations for this big database, we shifted to a cloud notebook. Fortunately, Kaggle hosted the whole database in their memory and we leveraged their storage and processing power

Insights on Data



This query exemplifies Yelp's ability to identify the most reviewed businesses, which can be crucial for businesses seeking to improve their reputation and customer engagement.

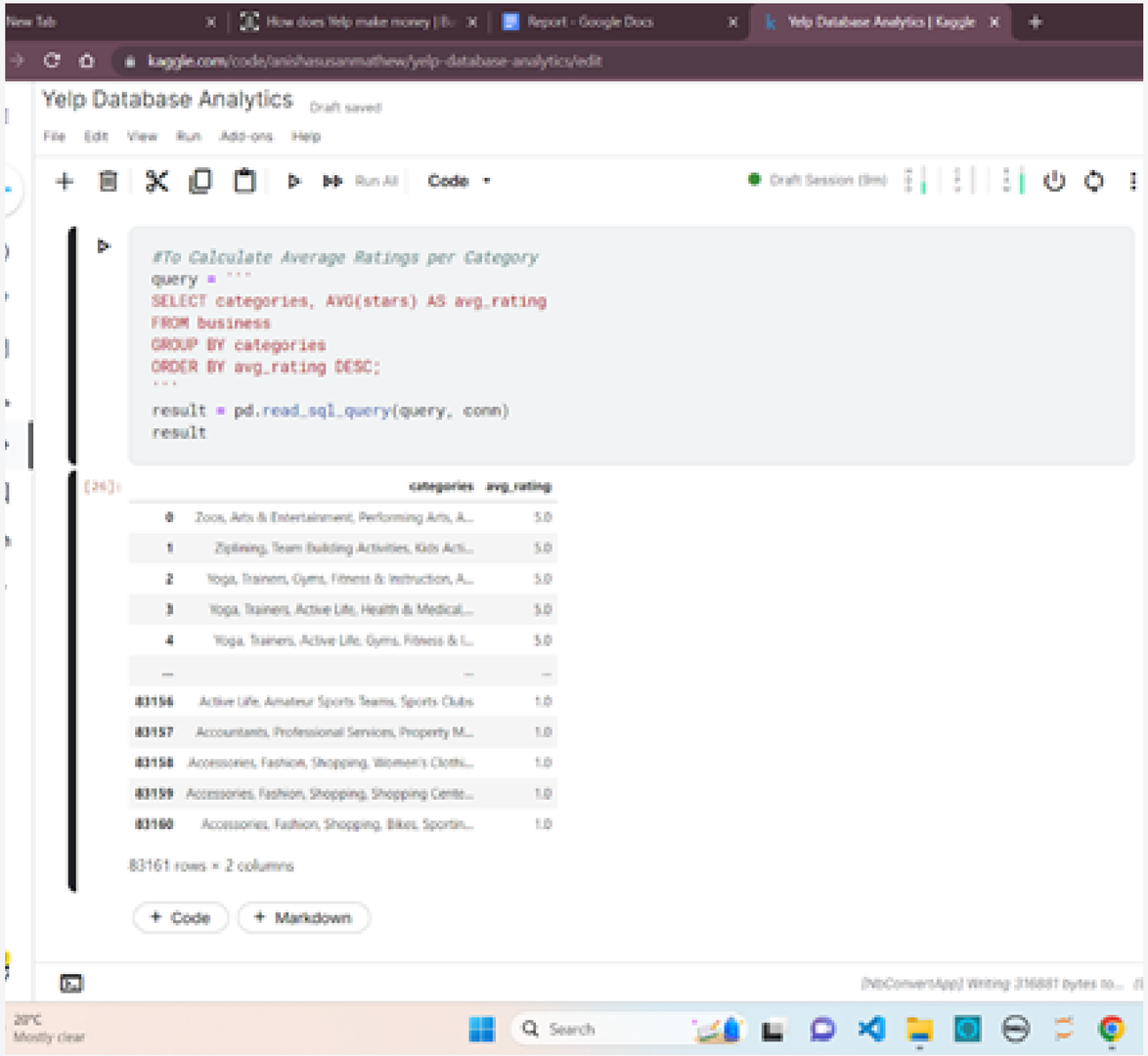
This is explained as follows :

1. Yelp's platform encourages user-generated reviews, resulting in businesses with high foot traffic accumulating numerous reviews, contributing to their visibility and popularity.

2. Yelp uses star ratings and sorting algorithms to display businesses with high average ratings and recent, engaging reviews at the top of search results.

3. Yelp's "Popular" and "Top" categories highlight businesses with substantial positive reviews, providing increased exposure and attracting potential customers.

4. Engaging with Yelp's community, including the Yelp Elite Squad, can further enhance a business's reputation and customer engagement on the platform.



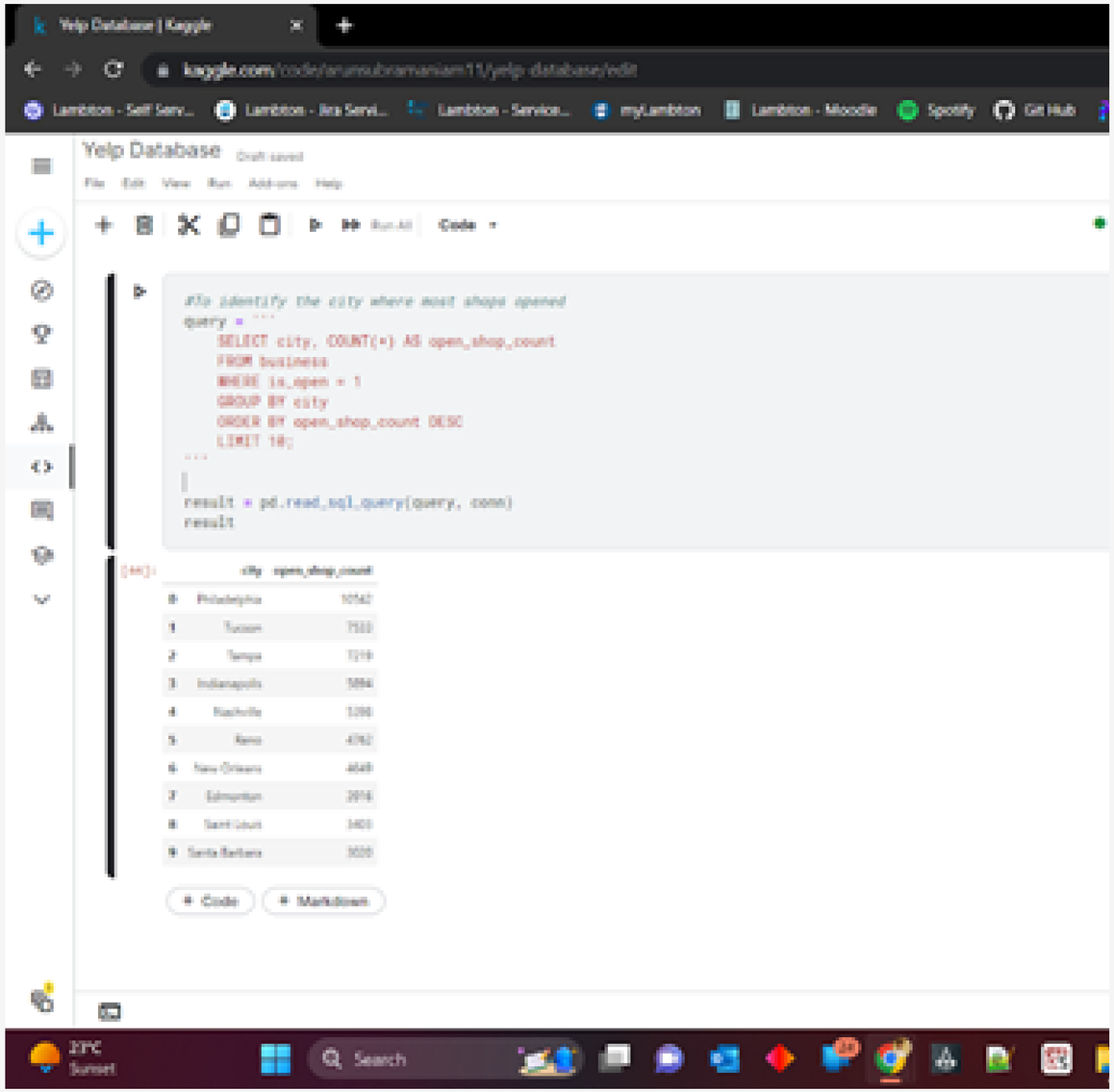
This query showcases Yelp's capacity to calculate and display the average rating for each business category, enabling users to identify the most popular and well-rated industries.

This is explained as follows:
1. Yelp calculates and displays the average rating for businesses in each specific industry or category based on user-generated reviews and star ratings.

2. The platform aggregates reviews for businesses within a category, allowing users to identify the most popular and well-rated industries.

3. Yelp's sorting and ranking algorithms prioritize higher-rated businesses within a category, making it easier for users to find the best-reviewed establishments.

4. Users can use filters and search options to refine their results and make informed decisions when choosing businesses, restaurants, or services in a particular category.



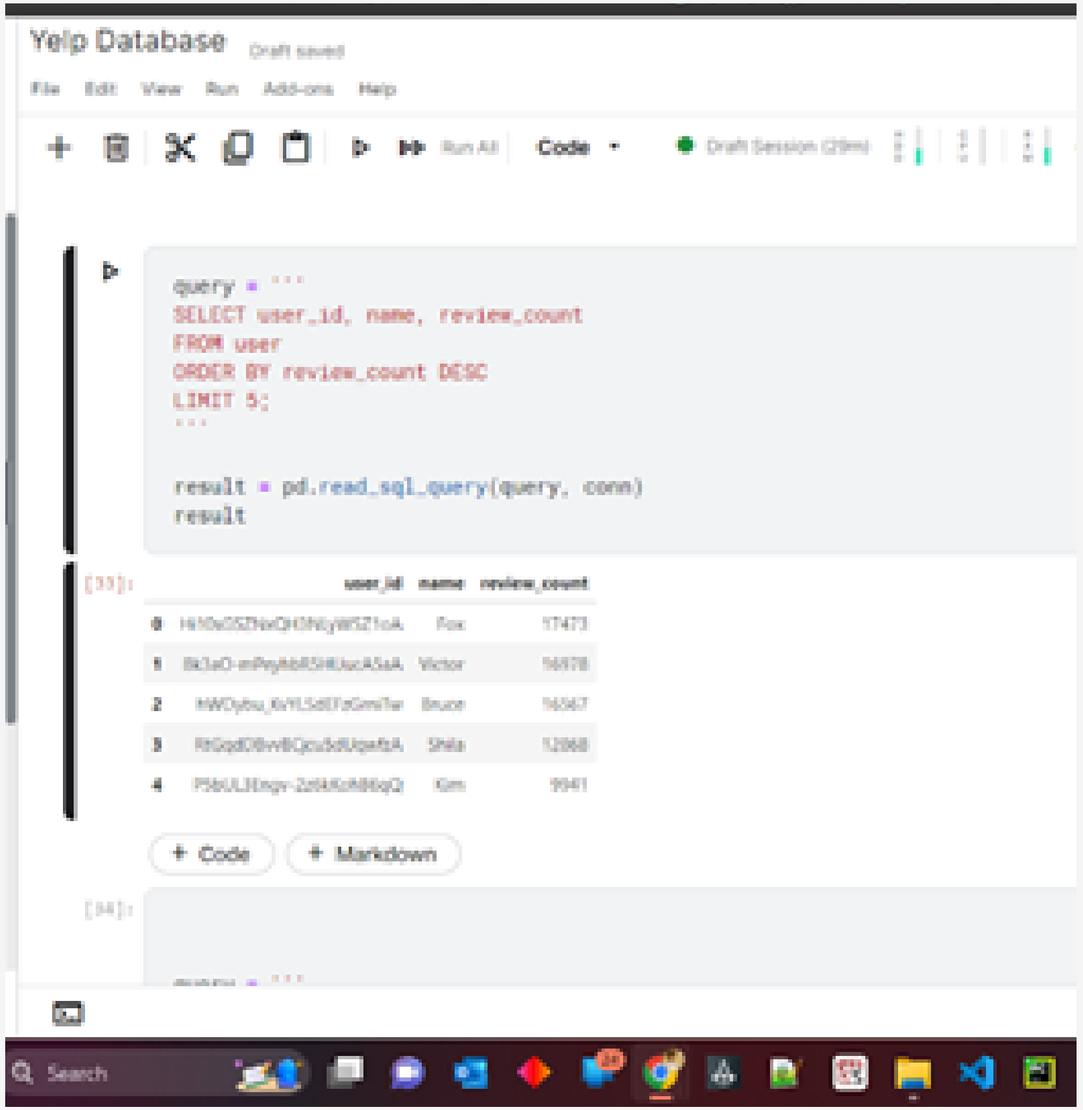
By identifying the city where most shops were opened, Yelp can use this information to improve its services and offerings in several ways:
Targeted Marketing: Yelp can focus its marketing efforts on the city where most shops were opened. This includes targeted advertising campaigns, sponsored listings, and promotions to attract more users and businesses to the platform in that city.

Expansion Planning: If a particular city shows significant growth in the number of shops being opened, Yelp can prioritize its expansion efforts in that area. This could involve increasing resources, customer support, and local partnerships to enhance the Yelp experience in that city.

Localized Content: Yelp can curate and promote localized content, including featured businesses, trending spots, and user-generated content from the city with the highest number of shop openings. This can help engage users in that city and encourage more reviews and interactions.

Business Insights: By analyzing the data from shops that have been opened in the identified city, Yelp can gain insights into the types of businesses that are thriving in that area. This information can be valuable for business owners and potential entrepreneurs looking to start new ventures.

Overall, identifying the city where most shops were opened can provide Yelp with valuable insights to tailor its strategies, offerings, and services to cater to the specific needs and preferences of users and businesses in that area.



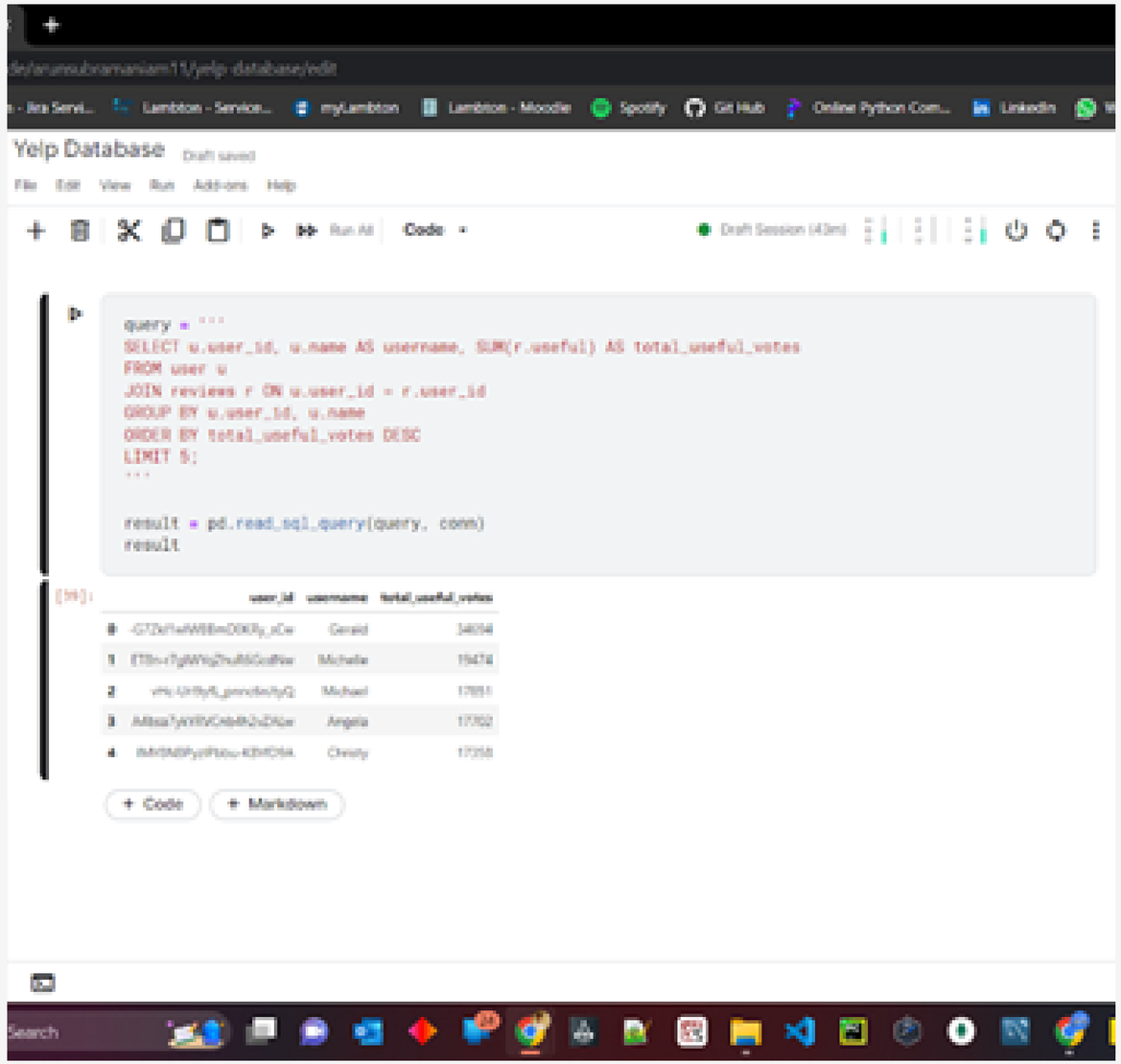
Yelp can use the insights gained from the top 5 users with the highest number of reviews on its platform to enhance various aspects of its service. Some are:

User Engagement: Yelp can create incentive programs or loyalty rewards for users who write a good number of reviews. This could encourage more users to actively engage with the platform and share their experiences.

Quality Control: One of the major problems yelp faced was fake reviews. Yelp can analyze the reviews from top users to identify patterns of high-quality content. This analysis can be used to develop guidelines and standards for other users to follow, improving the overall quality of reviews on the platform.

Social Media Promotion: Yelp can collaborate with top users to promote their reviews and experiences on social media channels. This cross-promotion can increase user engagement and attract new users to the platform.

Yelp can strengthen its community, enhance the quality of user-generated content, and provide more valuable information to users and businesses alike. This can lead to increased user engagement, improved business visibility, and a more vibrant and thriving Yelp ecosystem.



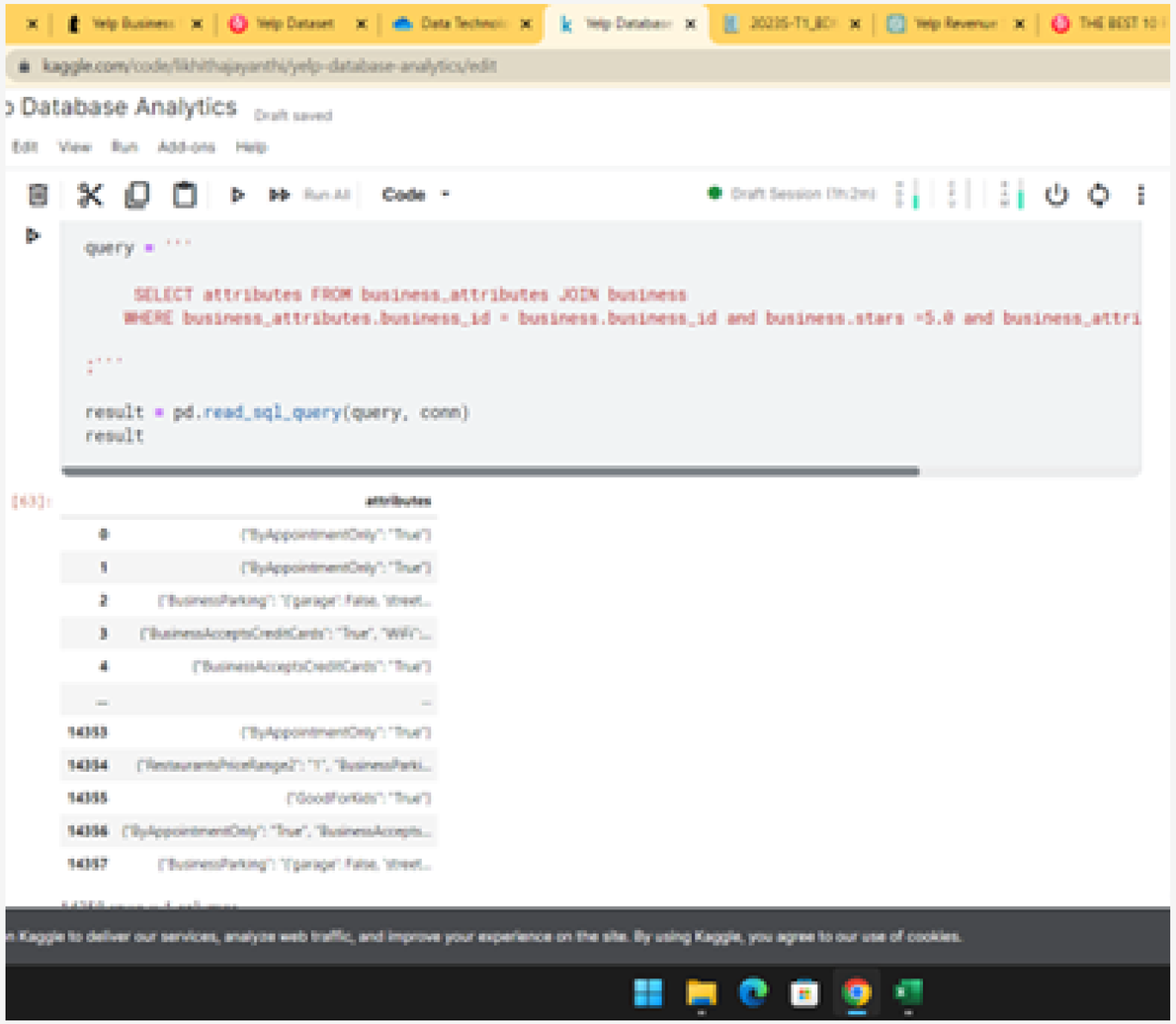
Yelp can capitalize on the valuable feedback and influence of the top 5 users with the most useful votes received for their reviews to improve its platform and user experience.

Moderation and Quality Control: Reviews from top users with useful votes can be given higher visibility and priority during content moderation. Yelp can use their reviews as exemplars of high-quality content and use them to set standards for other users.

Community Building: Yelp can organize meetups or virtual events for the top users to connect and share their experiences. These events can foster a sense of community and encourage more engagement among influential users.

Getting Suggestions: Yelp can designate these top users providing them with a direct channel to share their feedback, concerns, and suggestions with Yelp's team. This can foster a sense of ownership and partnership between Yelp and its influential users.

Yelp can foster a stronger user community, improve content quality, and enhance the overall platform experience. This, in turn, can attract more users and businesses, making Yelp an even more valuable resource for both consumers and local enterprises.



With the help of the above query Yelp can discover the amenities that have been top rated by customers and can suggest those amenities to other businesses which have low rating to pick up their businesses.

Preferred Amenities :

Attributes like "outdoor seating," "free Wi-Fi," and "good for groups" are popular among top-rated businesses.
Yelp can emphasize these amenities to encourage other businesses to adopt them and enhance customer satisfaction.

Pet-Friendly Environment :

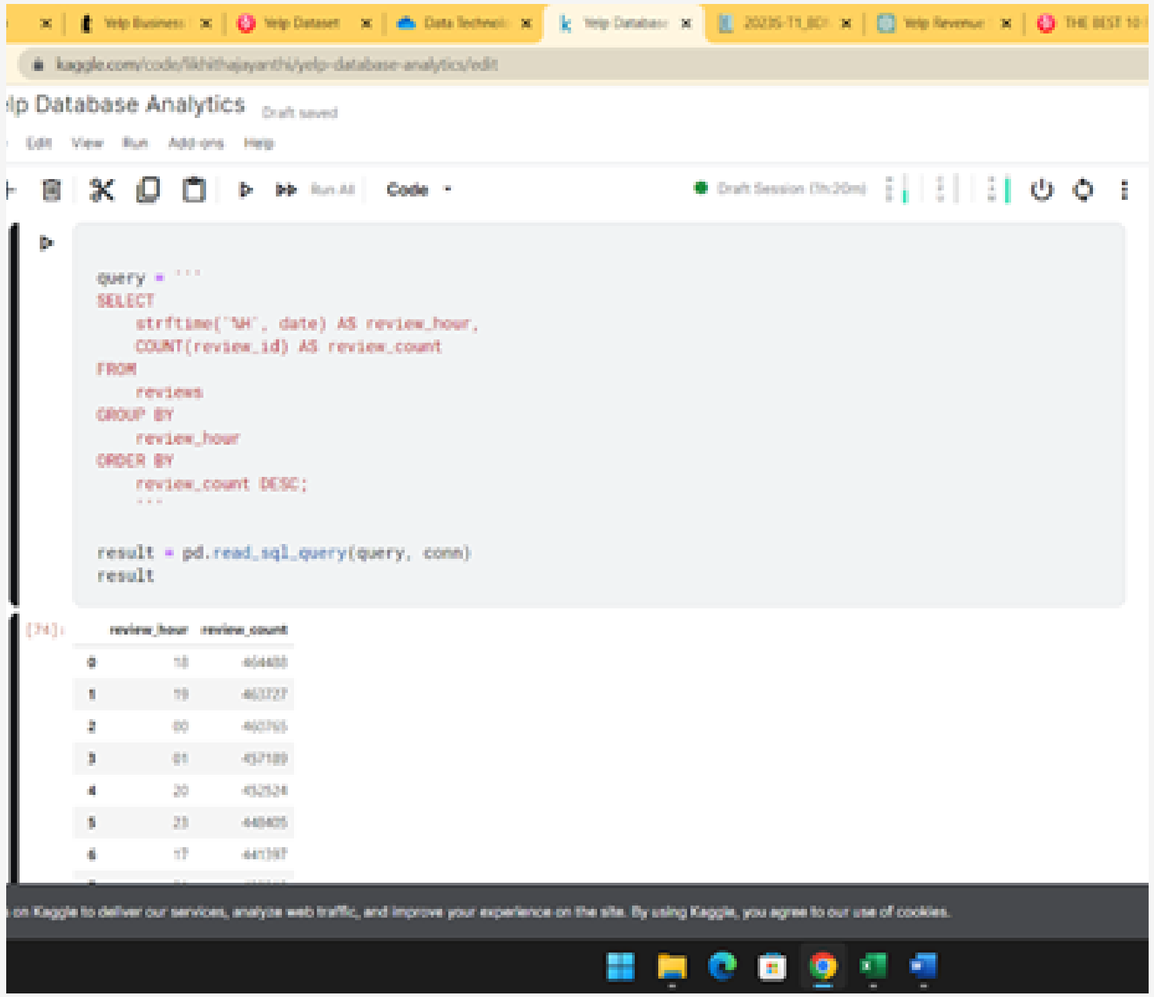
Top-rated businesses often have attributes like "pet-friendly," indicating a pet-welcoming environment.
Yelp can promote pet-friendly businesses, catering to users who wish to bring their pets along.

Exceptional Customer Service :

Attributes such as "friendly staff," "excellent service," and "knowledgeable staff" are associated with top-rated businesses.
Yelp can highlight the importance of outstanding customer service for businesses seeking higher ratings.

Unique and Attractive Ambiance :

Top-rated businesses are frequently praised for their "cozy ambiance," "stylish decor," and "unique atmosphere."
Yelp can encourage businesses to create distinct atmospheres to stand out and attract positive reviews.



with the query Yelp can Analyze the peak review times which can help Yelp identify when users are most active, allowing them to optimize the platform's content, push notifications, or marketing efforts during these hours to increase user engagement.

Busy Hours

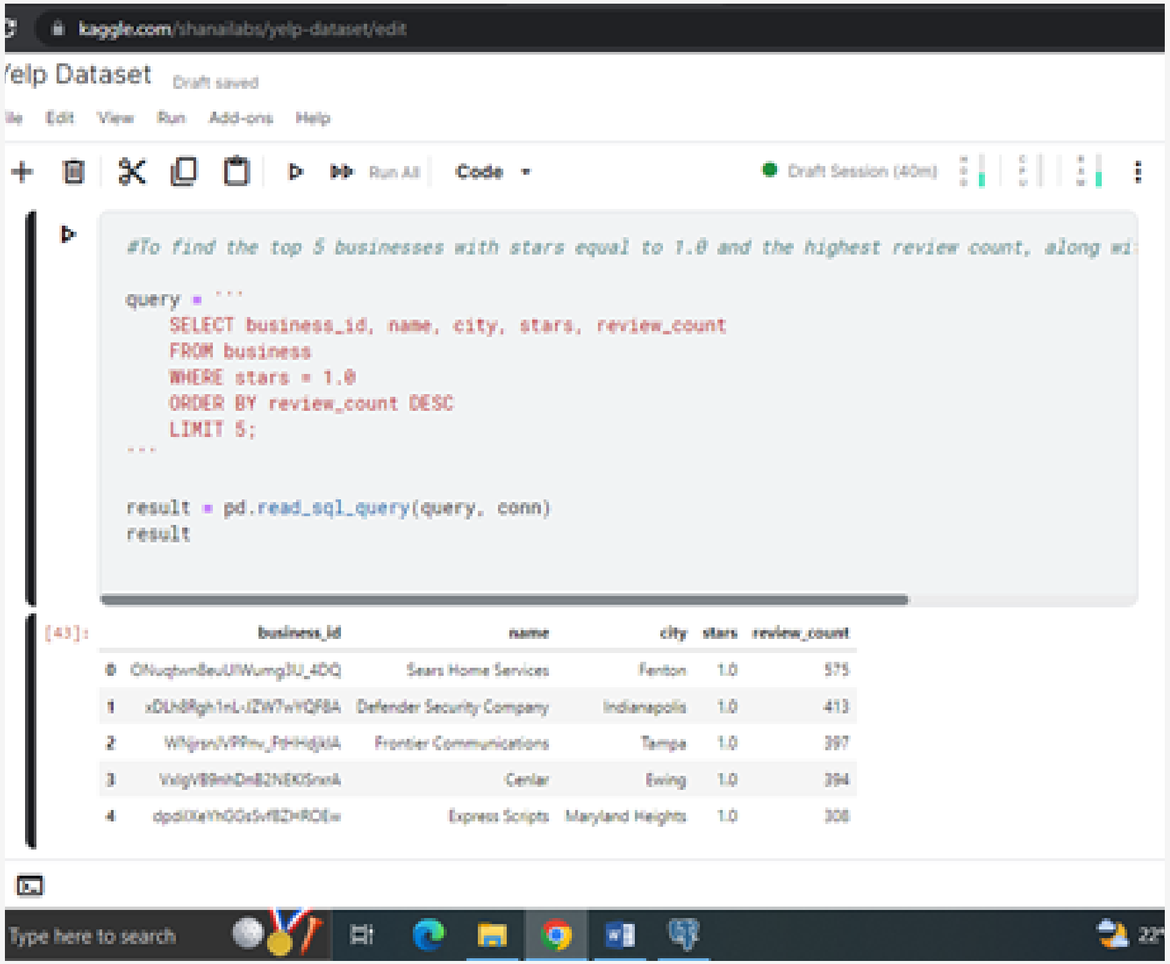
The query reveals peak review times during specific hours, such as 6 PM and 8 PM, suggesting high user activity during these periods.
Yelp can optimize content, push notifications, and marketing efforts during these busy hours to engage users effectively.

Lunchtime Reviews

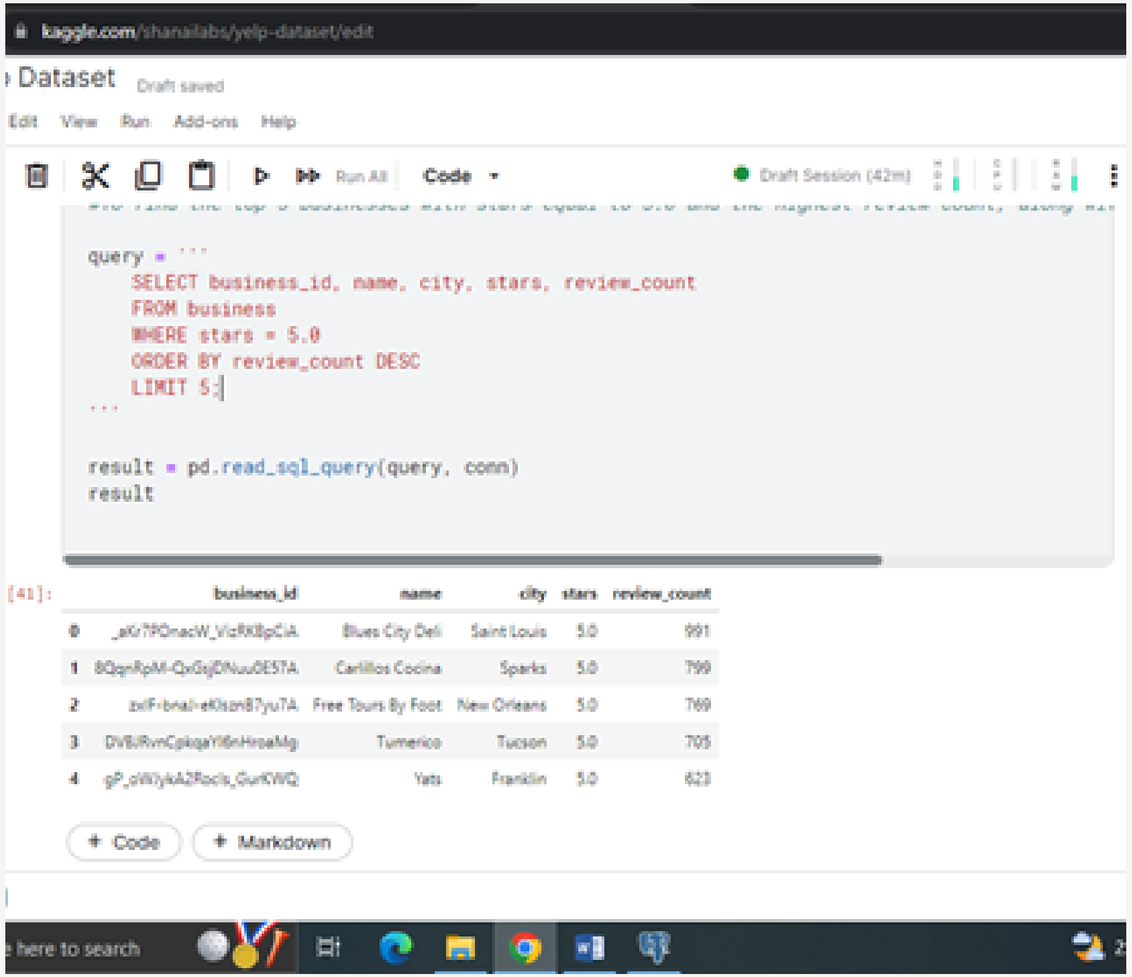
Peak review times around 12 PM indicate that users are actively posting reviews during lunchtime.
Yelp can use this information to promote lunch deals and specials to attract users during these hours.

Evening Engagement

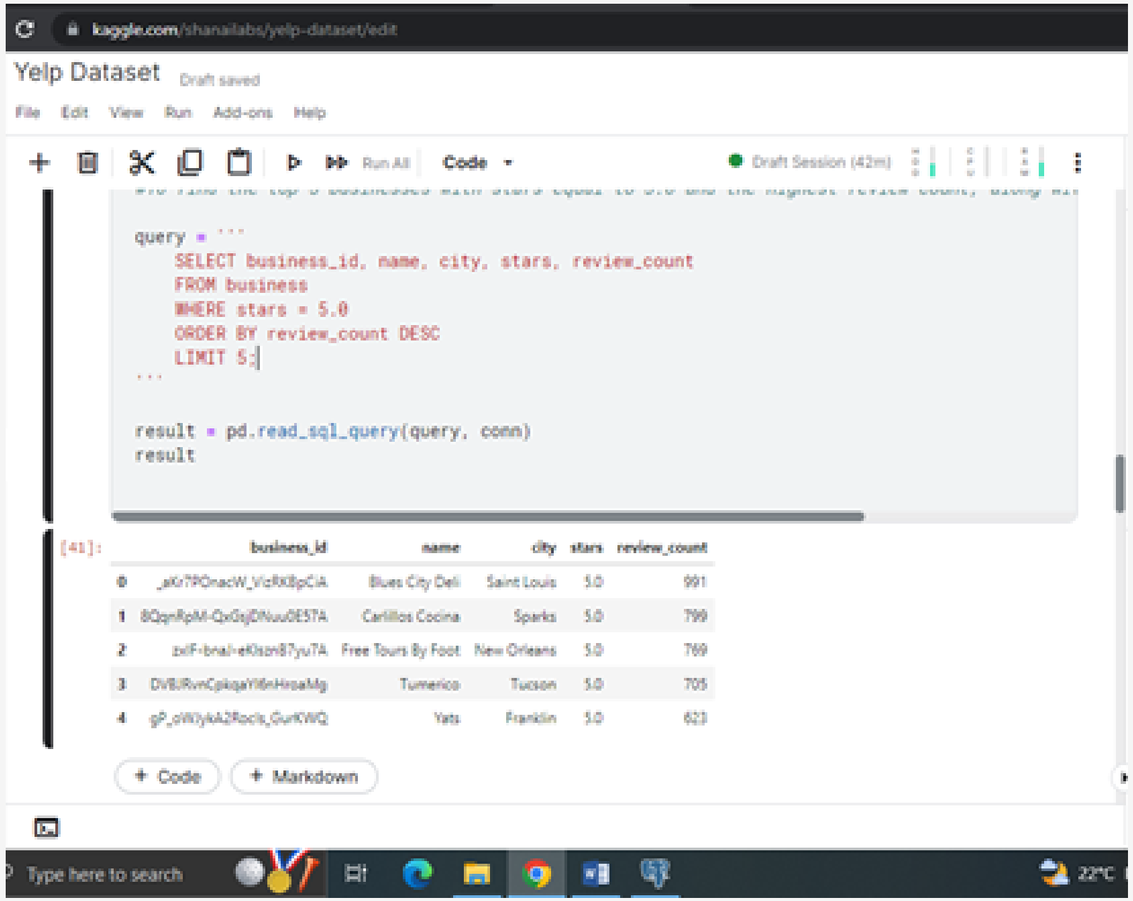
Reviews surge around 7 PM, indicating higher user engagement during the evening hours.
Yelp can consider featuring businesses with evening events or happy hours to capture this engaged audience.



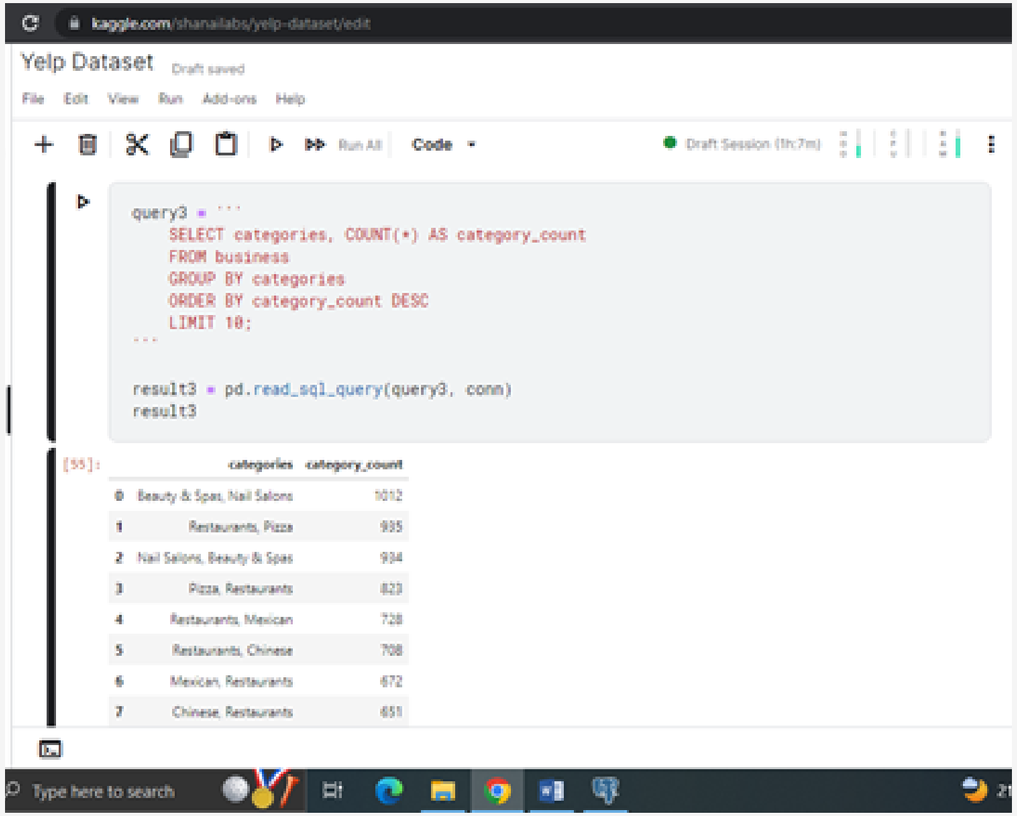
In this query we find the businesses with the minimum star rating with the most number of review counts to get the desired output.



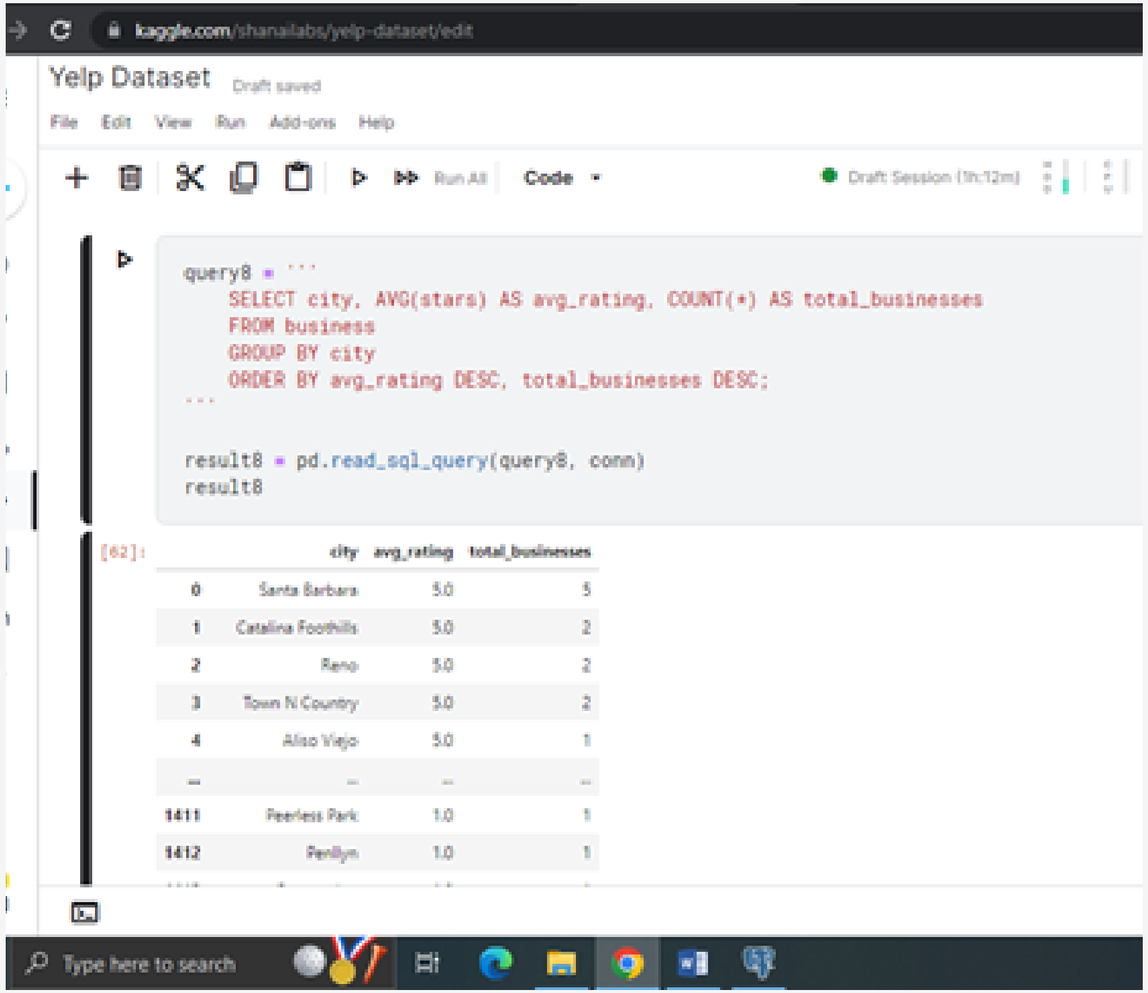
The below query is to find the top performing businesses with most 5 star ratings. This helps Yelp to identify the companies which have the most customer satisfaction, this can be used to give bonuses to the employees of the top business.



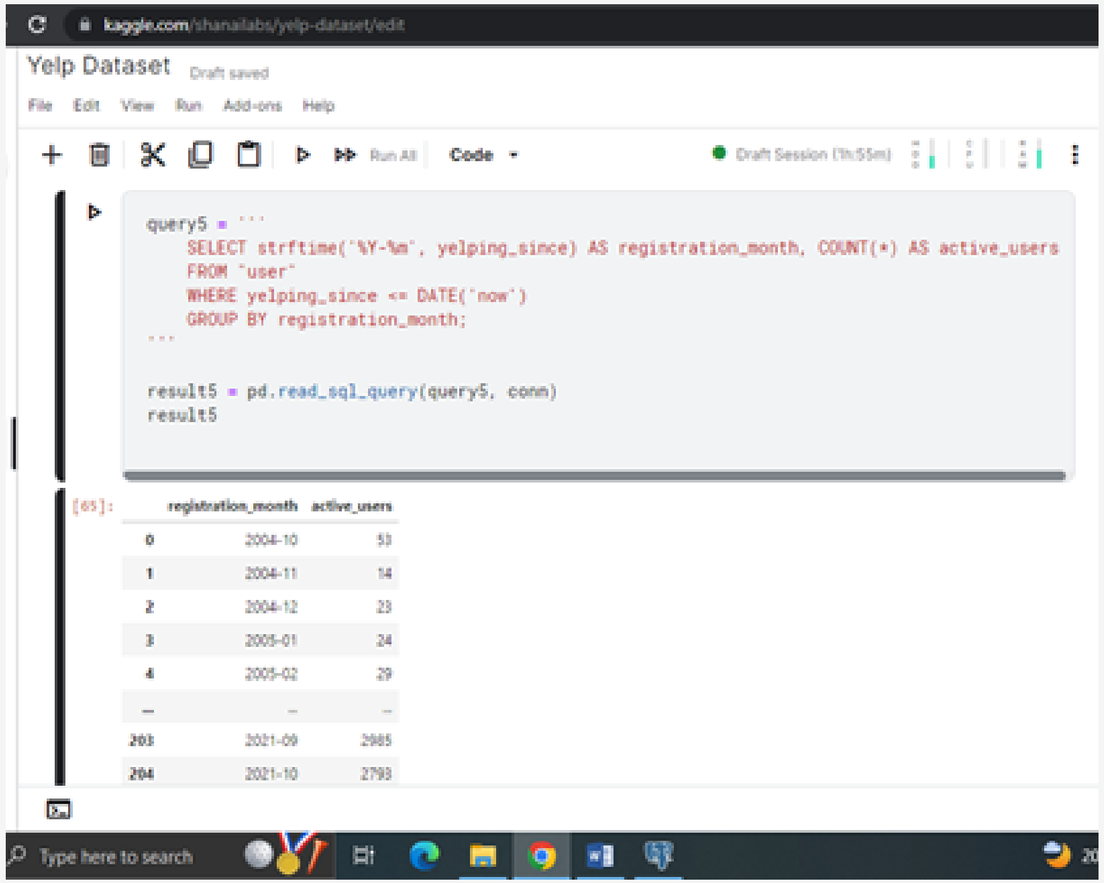
2.The below query is to find the top performing businesses with most 5 star ratings. This helps Yelp to identify the companies which have the most customer satisfaction, this can be used to give bonuses to the employees of the top business.



The query is to find the most number of Business Categories



The below query finds the location wise businesses count and their average rating. By this Yelp can find which location is best to start any new business or improve the business in the location where the average rating is low



5.The below query is to find the active users by each month. By this Yelp can identify active user growth over the year and make sure the user base increases in the future.

Final Submission PLAN

Overall, the final submission aims to provide a comprehensive and insightful analysis of the Yelp dataset, drawing meaningful conclusions and recommendations for businesses and users alike.

- We plan to further deepen the analysis and apply more advanced data analysis techniques and techlogies like Hadoop and Spark.
- Conducting sentiment analysis on reviews to understand the overall sentiments of users towards businesses.
- Analyzing geographic trends, such as the distribution of highly rated businesses across different locations.
- Creating interactive data visualizations to present the findings in a more user-friendly and informative way.
- Providing actionable recommendations to businesses based on the analysis to improve their services and ratings.

Conclusion

Through the SQL queries and understanding the business process of Yelp, we wanted to find out how we as analysts can contribute to their business growth. While this midpoint submission is just scratching the surface, we hope to produce grater insights using more big data technologies at the time of the final presentation.

Questions ?

¡ ⊬0⊬

THANK YOU !