

# BUSINESS ANALYSIS ON YELP DATASET

## CSE 511- MILESTONE PROJECT

MANSI NANDKAR (1233870562)

### INTRODUCTION:

Yelp is one of the most popular online review platform where users can leave star ratings and detailed reviews for businesses. It provides a one-stop local platform for consumers to discover, connect, and transact with local businesses of all sizes. It also serves as a critical resource for understanding consumer behaviour, identifying market trends, and driving data-driven business decisions. The primary goal of this analysis is to evaluate the key factors influencing business success and customer satisfaction. By leveraging advanced tools like PySpark, Hadoop, and Python, the project aims to uncover trends and actionable insights that can inform business strategies for operational improvement and market positioning.

### METHODOLOGY:

#### DATSETS:

The datasets are available in JSON format which are as follows:

NAME	INFO
USER	Contains user profiles, detailing reviewer information such as user ID, compliments received, average stars, review counts, social connections, and their yelping history.
TIP	Provides short text tips left by users for businesses, along with details such as user ID, business ID, date, and the number of compliments received for the tip.
REVIEW	Includes user reviews for businesses, containing star ratings, review text, usefulness votes, and the review's associated business and user IDs.
CHECKIN	Captures check-in data for businesses, showing timestamps of customer visits grouped by business ID and date.
BUSINESS	Contains information about businesses, including their names, locations, categories, star ratings, hours of operation, attributes (e.g., WiFi), and review counts.

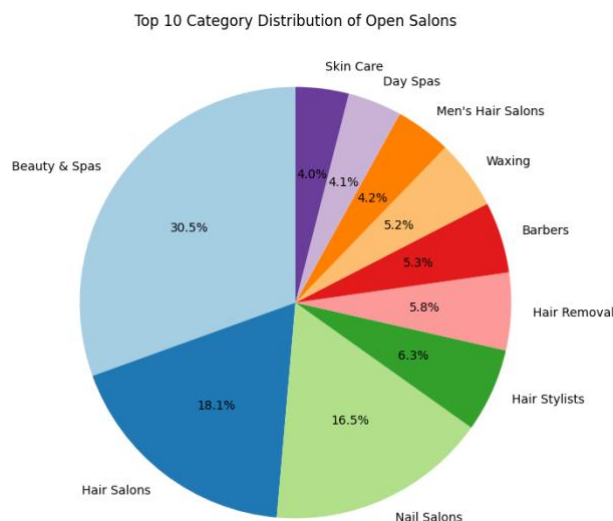
**TOOLS USED:** PySpark, Hadoop, Python

### IMPLEMENTATION:

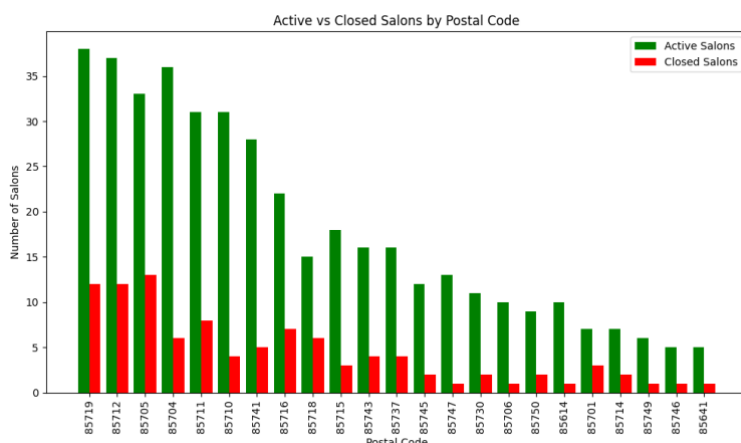
- Data Loading and Filtering:** The Yelp dataset was loaded in JSON files using PySpark and then I filtered the business dataset to include only businesses in Arizona (AZ). Further, I filtered to focus specifically on salons by searching for "Salon" in the categories which had 554 entries. I created temporary views of the filtered datasets to enable SQL queries for data transformation.
- Business Analysis:** The Key tasks included analysing the distribution of ratings, identifying popular categories associated with salons, and investigating attributes of high-performing businesses along with visualization to support the analysis.
- Data Visualization:** Various plotting techniques were employed to present the results of the analysis visually. Bar charts were used to depict rating distributions and geographical comparisons, making complex trends easier to interpret and communicate effectively.
- Insight Generation:** Insights derived from SQL queries and visualizations revealed critical trends in the salon industry in Arizona. Popular features such as exceptional service, convenient locations, and amenities like free WiFi emerged as factors contributing to high ratings. These findings offer actionable recommendations for improving customer satisfaction and operational success.

## FINDINGS & ANALYSIS:

1. The most famous category of salon was "Nail Salons, Beauty & Spas" which was around 105 salons out of which 30.5% is for Beauty & Spas in Arizona.



2. The location with 85719 had the highest number of customers i.e. 1916 with 38 active salons and 20 closed salons making up a total of 50 salons. The average number of customers in this location was 50.



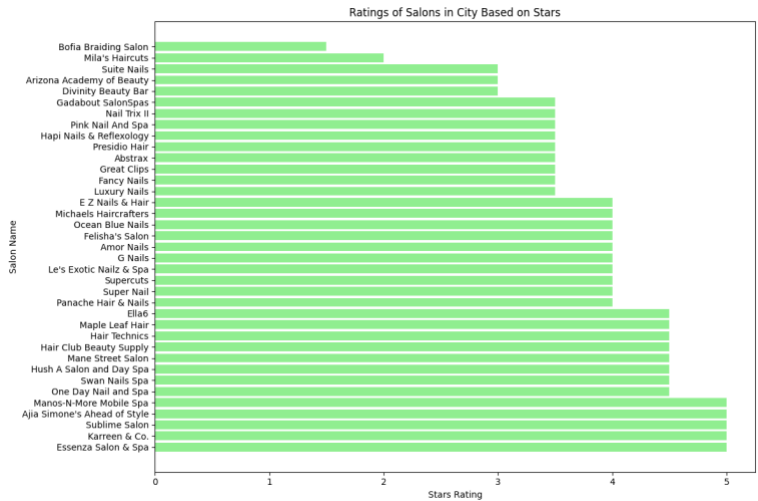
3. For a salon to be considered a good salon, the criteria included receiving high ratings from a user who has been active on Yelp for over 15 years, marked the review as useful and cool, and has a history of contributing to numerous businesses through their reviews on Yelp. This also indicating their credibility and engagement with the Yelp community.

business_id	busi_name	postal_code	stars	user_name	user_on_yelp_yrs	user_total_review
vwl6YkyTZVt035bU-...	Greentoes	85705	5.0	Margot	16	922
_ijPwDwnqcvrNfmNN...	Styling Etc.	85747	5.0	Betsy	17	851
DPEh8VvJ7WgJpft2h...	Belle Starr Salon	85745	5.0	Amanda	16	444
vnEpznJh2EGVIFE10...	Salon Salon	85701	5.0	Amanda	16	444
HJq6nfiPa6_2FhMX-...	Red Nails	85711	5.0	Debbie	16	162
_10KD3jnwKVfFrq6d...	The Nail Lounge O...	85719	5.0	Veronica	16	129

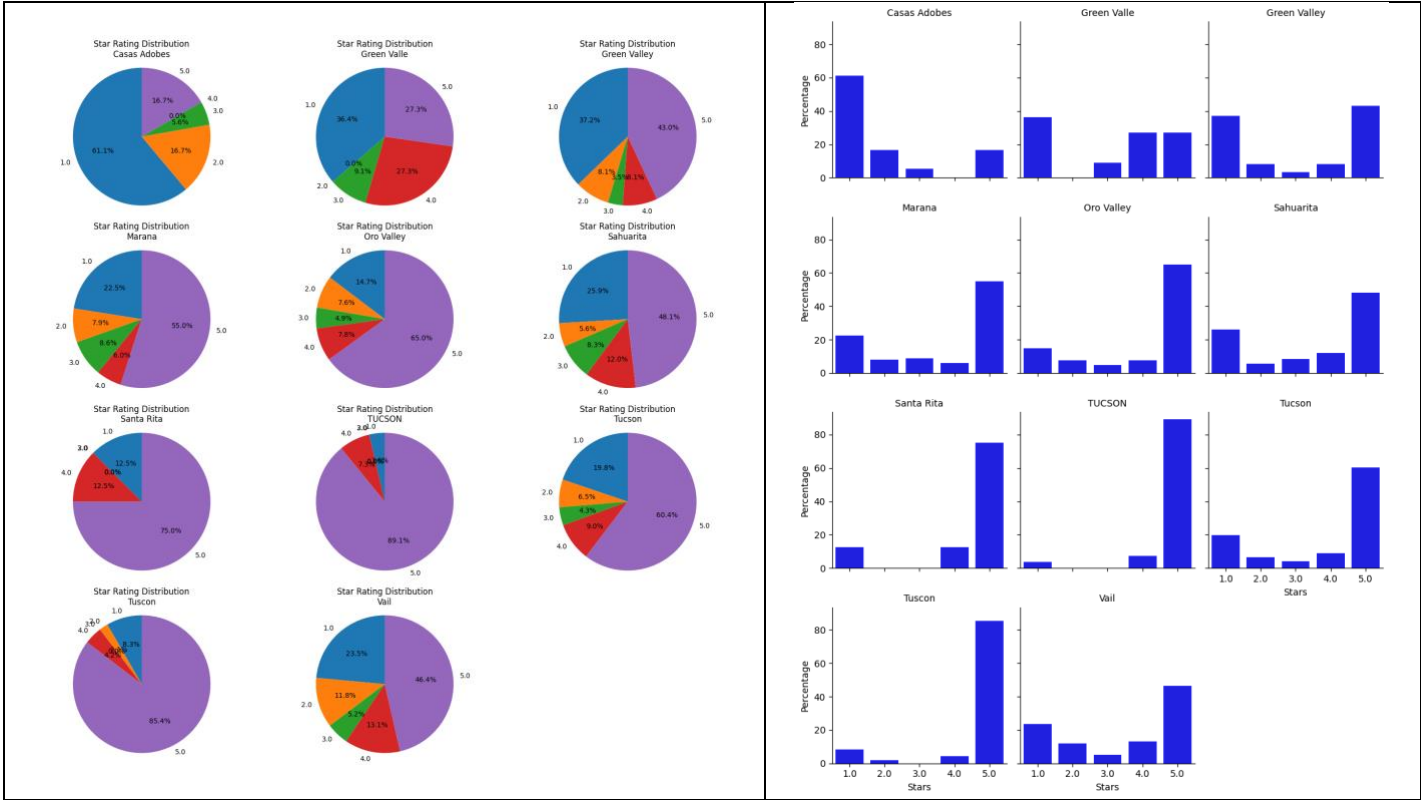
4. For postal code 85704, "Lovely Nails" has a high rating of 4.5 stars and belongs to a postal code with 42 total salons, 36 of which are active. Similarly, 85712, featuring "Maple Leaf Hair" and "Gadabout Salon Spas" with ratings of 4.5 and 3.5 stars, has 49 total salons, of which 37 are active. This also shows that high-rated salons are often found in postal codes with a higher proportion of active salons with specific categories such as "Nail Salons," "Day Spas," and "Skin Care."

postal_code	busi_name	stars	categories	compliments
85704	Lovely Nails	4.5	Nail Salons, Waxi...	3
85712	Maple Leaf Hair	4.5	Beauty & Spas, Ta...	2
85706	Hollywood Nails &...	2.5	Massage Therapy, ...	2
85711	Ulta Beauty	3.5	Skin Care, Hair S...	2
85712	Gadabout SalonSpas	3.5	Day Spas, Beauty ...	2

5. The plot displays the star ratings of salons in a specific city (For Eg: postal code 85712 for this graph). Each salon's rating is represented by a horizontal bar, with the length corresponding to its star rating. This allows easy comparison of salon ratings, highlighting those with the highest ratings. The chart provides insights into the overall quality of salons in the selected postal code area.



6. The plot provides a city-wise comparison of salon ratings, displaying the percentage distribution of each star rating across cities. It helps visualize how salons in different cities are rated, with each city having its own bar plot. This allows for easy identification of cities with higher concentrations of specific star ratings. The visualization also aids in comparing overall customer satisfaction across multiple cities.



## CONCLUSIONS & RECOMMENDATIONS:

1. **Focus on Core Services:** The analysis revealed that hair styling and general beauty services dominate as the most popular categories among highly-rated salons. Salons focusing on these core offerings are more likely to succeed, as they align with customer preferences. For new or expanding salons, prioritizing these services while also offering niche options like nail care or massage therapy could help attract a diverse customer base.
2. **Accessibility & Inclusivity Matters:** Key attributes of successful salons include being wheelchair accessible and accommodating families, reflecting the importance of inclusivity in attracting a broader range of clients. Salons that invest in such features can differentiate themselves in competitive markets by ensuring convenience and comfort for all customers.
3. **Prioritizing Customer Service:** High-rated salons consistently emphasize exceptional customer service, skilled staff, and a welcoming atmosphere. Positive reviews often mention these factors as pivotal to customer satisfaction. Investing in staff training and maintaining a friendly and professional environment can significantly enhance customer loyalty and salon reputation.
4. **Managing Online Reputation:** With an average of 25.7 reviews per salon, a strong online presence emerged as a critical factor for success. Salons with high ratings are often supported by credible reviews from experienced users. To leverage this, salons should encourage satisfied customers to leave positive reviews, actively engage with feedback, and address any negative reviews promptly to build trust and attract new clients.
5. **Strategic Location Selection:** The geographical analysis highlighted Phoenix and Scottsdale as key hubs for successful salons, with specific zip codes like 85704 and 85712 showing high performance in terms of active salons and customer engagement. Businesses planning to expand should prioritize these high-demand areas, as they offer greater potential for visibility and profitability due to concentrated customer interest.

By focusing on these areas, salon owners and managers in Arizona can improve their chances of success in a competitive but thriving market.