COFFEE KING

YELP DATA ANALYSIS

COFFEE KING IS A NEW STARTUP COFFEE COMPANY PROVIDING A UNIQUE AND NOVEL EXPERIENCE TO THEIR CUSTOMERS. THEY WANT TO APPEAL TO A WIDE VARIETY OF CLIENTELE. YOU WILL USE YELP REVIEWS AND USER DATA TO PROVIDE INSIGHTS TO COFFEE KING FOR THINGS SUCH AS LOCATION SELECTION AND/OR HOURS OF OPERATION (AND OTHER QUESTIONS?)

TABLE OF CONTENTS

SECTION 1: REVIEW QUESTION

GOAL

- Coffee King is a new startup coffee company providing a unique and novel experience to their customers. They want to appeal to a wide variety of clientele. You will use Yelp reviews and user data to provide insights to Coffee King for things such as location selection and/or hours of operation (and other questions?)

- Provide insights to Coffee King to assist in determining:

    - possible locations

    - best hours for operation

    - characteristics of successful businesses

    - the largest competitors

    - And other found insights

SECTION 2: HYPOTHESES

INITIAL QUESTIONS AND HYPOTHESES

Questions

- What relationship do the star ratings have to do with the businesses being open or closed?

- Do the reviews of the businesses assist in determining the star rating of the businesses?

- What, if anything, do reviews say about the businesses with high star ratings and number of reviews that make the customer continue to frequent the establishment.

Hypotheses

- The location of the business and the name will drive the customers to the establishment.

- Reviews will have a relationship with the business status (open / close).

- Value of the text (negative to positive) will have a positive relationship with the business' star rating.

SECTION 3: APPROACH

STEPS OF DATA ANALYSIS


```
1  ## looking at the number each category to determine where the null values are
2  SELECT COUNT(business_id) AS num_bus_id, COUNT(name) AS num_name, COUNT(stars) AS num_stars, COUNT(postal_code) AS
   num_postcodes,
3  COUNT(latitude) AS num_lat, COUNT(longitude) AS num_long, COUNT(state) AS num_state, COUNT(address) AS num_address,
4  COUNT(city) AS num_city
5  FROM `carla-project-1.business.business`
```

Press Alt+F1 for Accessibility Options.

Query results   SAVE RESULTS ▾   EXPLORE DATA ▾

JOB INFORMATION   RESULTS   JSON   EXECUTION DETAILS   EXECUTION GRAPH PREVIEW

Row	num_bus_id	num_name	num_stars	num_postcodes	num_lat	num_long	num_state	num_address	num_city
1	150302	150302	150302	150302	150302	150302	150302	150302	150302

Retrieved data from YELP website.
https://www.yelp.com/dataset/download

Uploaded data sets and data tables using Google Big Query

Reviewed the data to determine quality, number of business and relevant categories

*Unsaved query ▾ ✕   ➕

▶ RUN   💾 SAVE ▾   👤 SHARE ▾   🕐 SCHEDULE ▾   ⚙ MORE ▾

```
1  ## determining the number of business per category
2  SELECT categories, COUNT(business_id) AS num_businesses
3  FROM `carla-project-1.business.business`
4  GROUP BY categories
5  ORDER BY num_businesses DESC
6
7
```

Query results

JOB INFORMATION   RESULTS   JSON   EXECUTION DETAILS

Row	categories	num_businesses
1	Beauty & Spas, Nail Salons	1012
2	Restaurants, Pizza	935
3	Nail Salons, Beauty & Spas	934
4	Pizza, Restaurants	821
5	Restaurants, Mexican	727
6	Restaurants, Chinese	708
7	Mexican, Restaurants	672
8	Chinese, Restaurants	651
9	Food, Coffee & Tea	508
10	Beauty & Spas, Hair Salons	493
11	Hair Salons, Beauty & Spas	480
12	Coffee & Tea, Food	473
13	Automotive, Auto Repair	361

Determined the most relevant data from the Coffee & Tea, Food category

Created a new data set and table using only the data relevant to the Coffee & Tea, Food Category

STEPS OF DATA ANALYSIS – NEW DATA SET / DATA TABLE

- Narrowed the focus on the 473 businesses in the Coffee & Tea, Food category

- 333 Open businesses

- 140 Closed businesses

CURRENT BUSINESS STATUS

Closed
140
30%

Open
333
70%

coffee_tea_food - 12/10/22, 8:23 PM

ABC	business_id
ABC	categories
🌐	city
ABC	DogsAllowed
ABC	DriveThru
ABC	Friday
123	is_open
123	latitude
123	longitude
ABC	Monday
ABC	name
×✓	Open24Hours
123	review_count
ABC	Saturday
123	stars
ABC	state
ABC	Sunday
ABC	Thursday
ABC	Tuesday
ABC	Wednesday

STEPS OF DATA ANALYSIS – LOCATIONS OF THE BUSINESSES

Determine the location of the relevant businesses

Determine where possible gaps in the market

SECTION 4: TECHNICAL CHALLENGES

DATA LIMITATIONS

All data has limitations. Some that is found here relates to no data from customers in the Mountain Standard Time Zone.

Most of the customers are on the East Coast.

Only 13 states in the United States and 1 provenance in Canada are represented in the data.
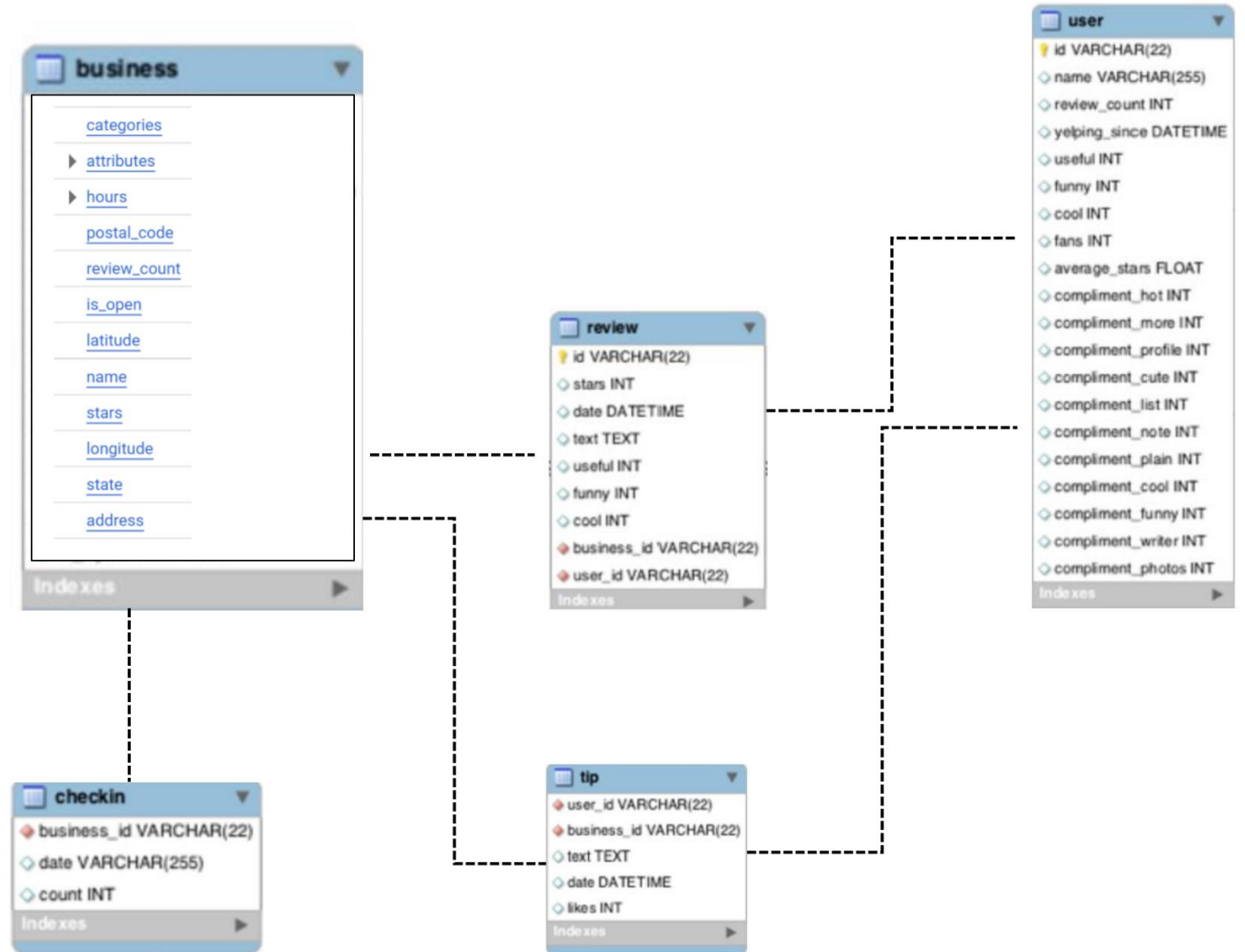
COFFEE KING

SECTION 5: ENTITY RELATIONSHIP DIAGRAM (ERD)

ERD

YELP DATA SETS DOWNLOADED FROM THE YELP WEBSITE AND THEIR RELATIONSHIP

SECTION 6: INITIAL FINDINGS

DESCRIPTIVE STATISTICS

```
## descriptive stats for Coffee & Tea, Food category
SELECT is_open, COUNT(business_id) AS num_of_business, MAX(review_count) AS max_rev_count, MIN
 (review_count) AS min_rev_count, Min(stars) AS min_stars, MAX(stars) AS max_stars, AVG(stars)
 AS avg_stars
FROM `carla-project-1.Coffee_Tea_Food.coffee_tea_food`
GROUP BY is_open
```

- Minimum and maximum number of stars

- Review count

- Average star rating grouped by the business being open or closed

Row	is_open	num_of_busines	max_rev_count	min_rev_count	min_stars	max_stars	avg_stars
1	0	140	133	5	1.5	5.0	3.65714285...
2	1	333	998	5	1.5	5.0	3.46846846...

- Open businesses had a more reviews
- Both open and closed businesses had the same minimum and maximum star ratings.
- Closed business had a higher average star ratings

DESCRIPTIVE STATISTICS

- Open vs Closed Businesses
  - Drive Thru
  - Allow dogs
  - Wifi
 - No Wifi

Characteristics

Stacked bar chart showing Open vs Closed values:
- Drive Thru: Open 36, Closed 5
- Oepn allow dogs: Open 59, Closed 6
- Free wi-fi: Open 180, Closed 112
- No wi-fi: Open 12, Closed 16

Legend: ■ Open ■ Closed
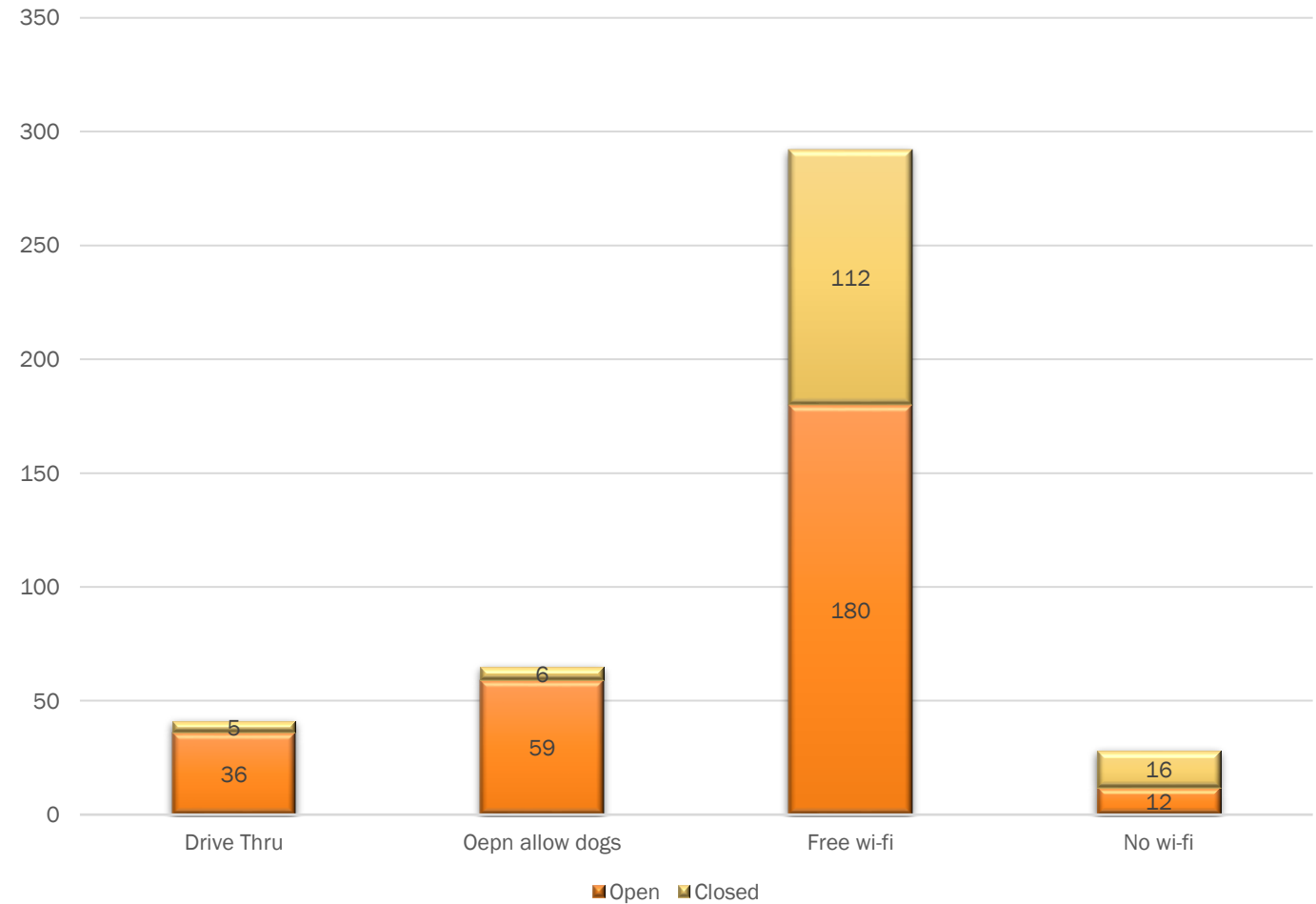
DESCRIPTIVE STATISTICS

- Importance of WiFi to a business

- More business with WiFi have reviews

- The average rating is higher for open businesses without WiFi

  - There are fewer businesses that do not have WiFi, which could be the reason for this occurrence (lower denominator when finding the average)

WiFi

Average Stars
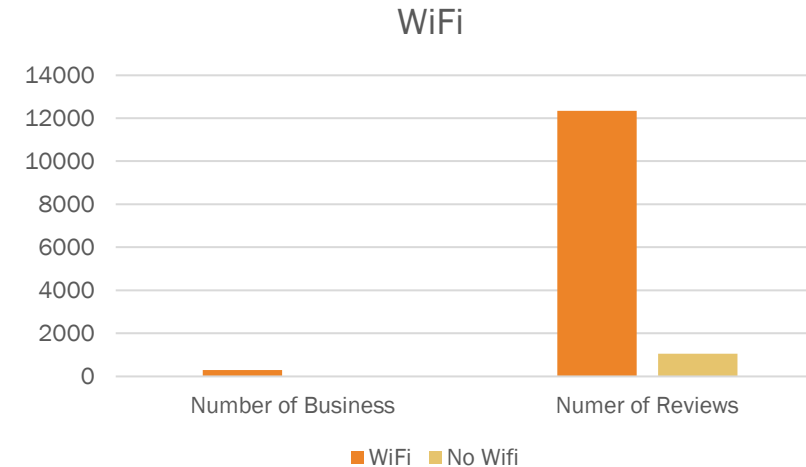
DESCRIPTIVE STATISTICS

- Number of business by name

- Starbucks has the most businesses in this category
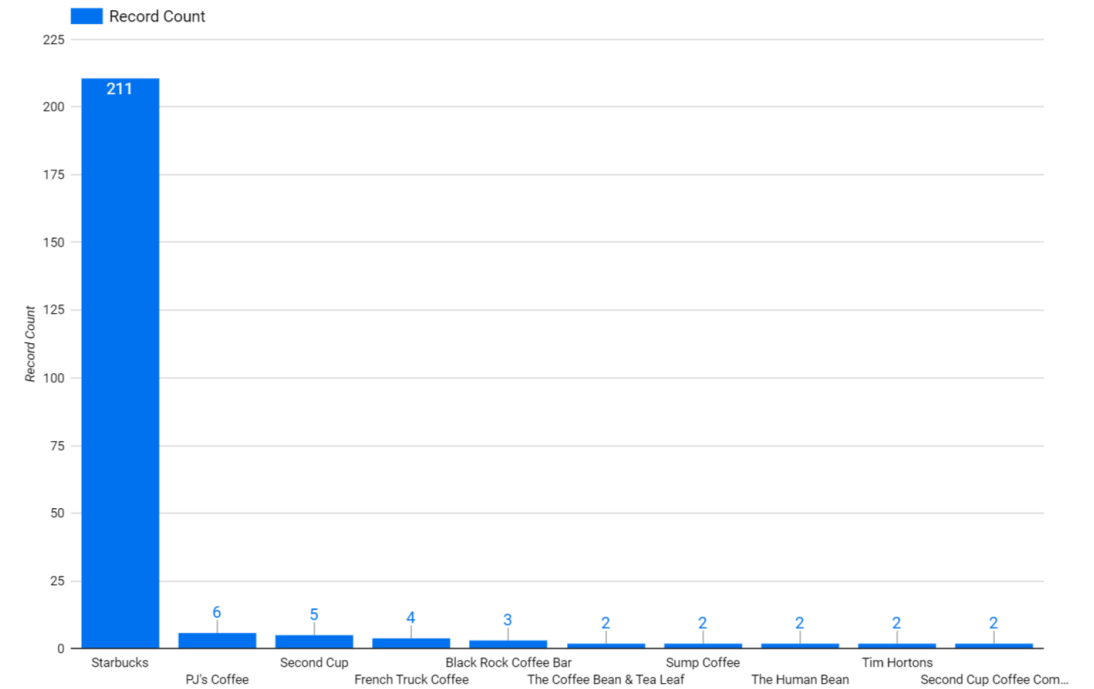
SECTION 7: DEEPER ANALYSIS

DEEPER ANALYSIS OF STARBUCKS

	st...	is_open ▾
1.	FL	40
2.	PA	34
3.	MO	21
4.	AZ	20
5.	TN	19
6.	AB	16
7.	IN	15
8.	NV	13
9.	NJ	8
10.	DE	6
11.	ID	6
12.	LA	6
13.	IL	4
14.	CA	3

1 - 14 / 14 ‹ ›

- Number of Starbucks by State
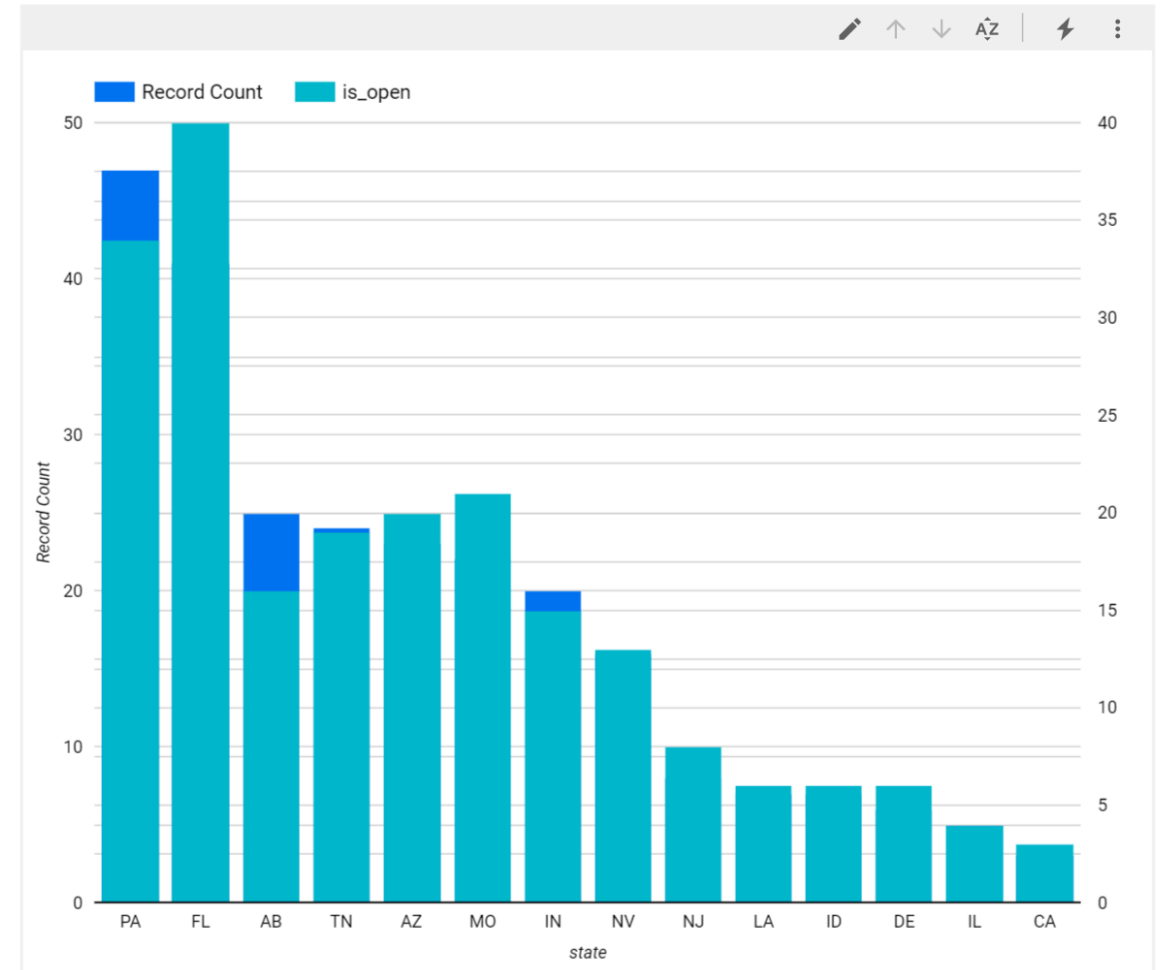
DEEPER ANALYSIS OF STAR RATINGS

- Reviewed the number of stars given by category

- Low rating: stars <=2

- Average rating: 2.5<= star rating <=3.5

- High rating: stars >=4

Number of Star Ratings by Value
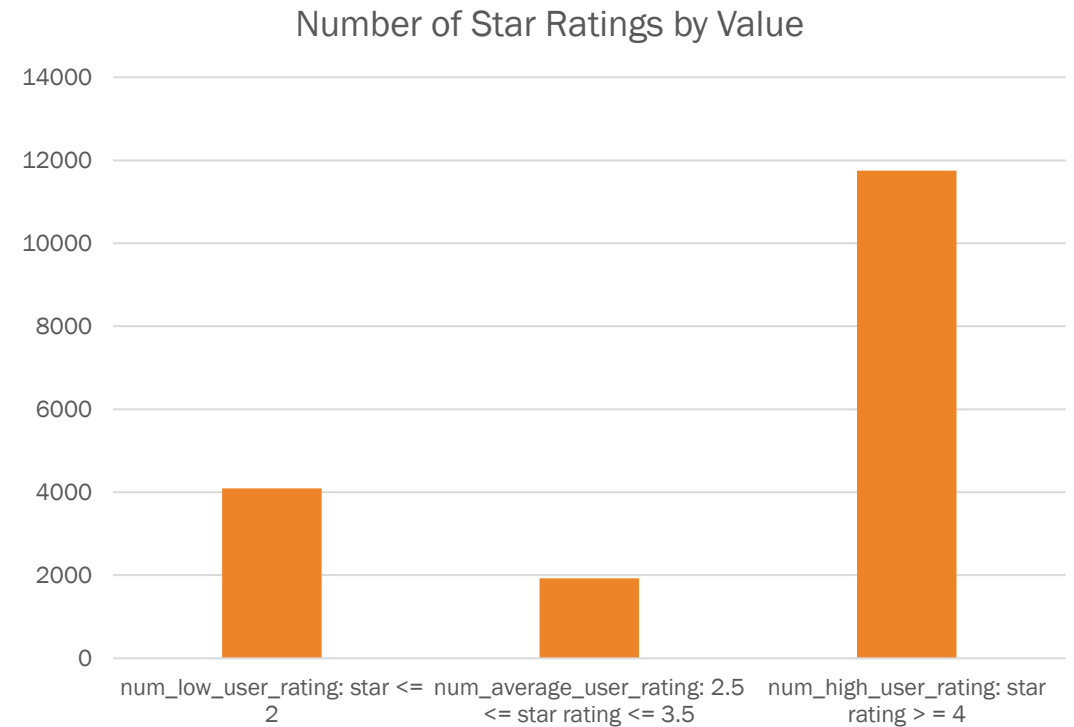
DEEPER ANALYSIS OF TEXT COMMENTS

- Words were reviewed in the comments.

  - Low value words: Hate, Worst, Terrible
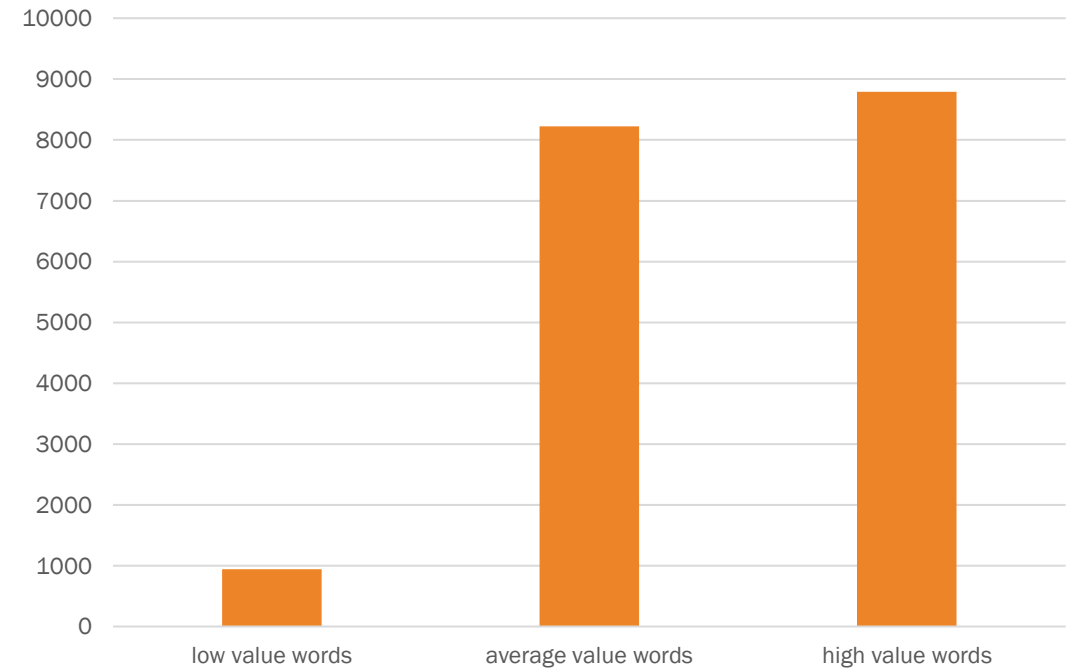
  - Average value words: Good, Ok, Average

 - High value words: Great, Best, Love, Excellent

Text Analysis - Word Value

	low value words	average value words	high value words

CORRELATION

- There is not a definitive relationship between the word value and the star rating.

- The **WORD VALUE** is a positive relationship. As the value of the word increases more positive words are in the comments.

- The **STAR RATING** decreases and then increases again, inferring that customers are more inclined to leave a star rating when they are either very dissatisfied or very satisfied.

RELATIONSHIP BETWEEN WORD VALUE AND STAR RATING

▨ Word Value ▨ Star Rating

	Low	Average	High
Word Value	941	8225	8794
Star Rating	4093	1927	11748
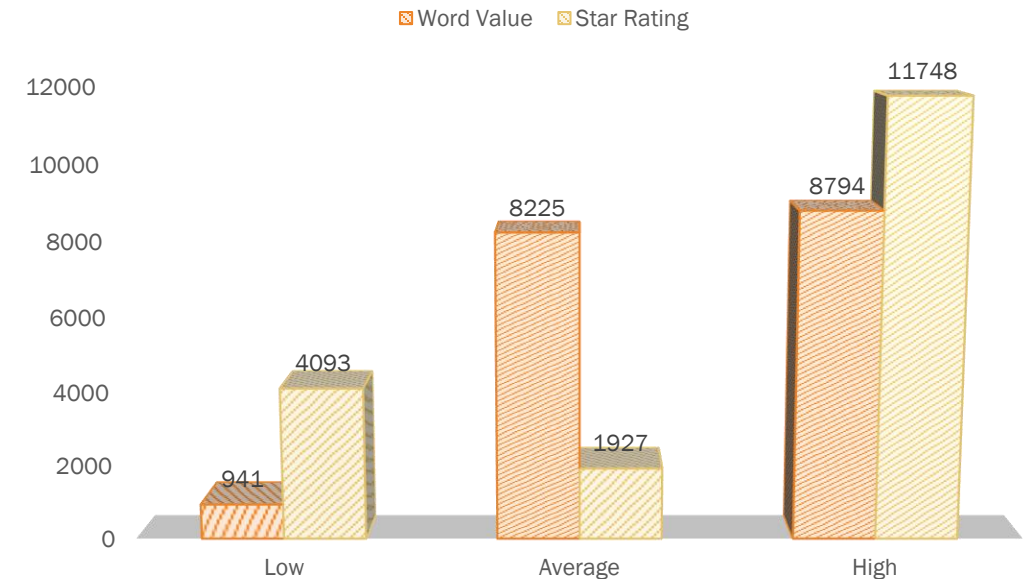
SECTION 8: HYPOTHESIS RESULTS / FINAL FINDINGS

FINDINGS

- The highest average ratings went for the closed businesses. However, this was also because they had fewer reviews therefore a lower denominator when finding the average. The more reviews given the more this will level out.
- The closing of the business seemed to be more for timing (2020) than to not being liked by customers. This should be investigated further. The closing times here were really based on when the last reviews or ratings were submitted. However, this could be based on many things.
- Only 4 closed businesses had below 3 stars (they all had 2.5) 2 businesses had 3.0 and the remaining had 3.5 and higher. Again, with a lower number of reviews this means the denominator is smaller when finding the average thereby affecting the overall average.
- More analysis into the business that closed is needed to determine when they closed and investigate the neighborhoods around them and if a Starbucks or a larger chain came into their neighborhood which may have driven them out.

COFFEE KING®

HYPOTHESES RESULTS

- **The initial hypothesis of Starbucks being the biggest competitor was proven.**
- Starbucks was the biggest competitor with 211 businesses out of 333 total business in the Coffee & Tea, Food category.
- Looked more closely at the Starbucks. Found the number of outlets by states (open and closed) to help the client determine where they could open a branch without the established competition. This could allow the client to establish themselves in the market and gain a solid customer base where Starbucks would not be able to undermine.
- Need to look at more data about the location of the Starbucks. Yelp users only reviewed this category in 14 states. Other data should be investigated.
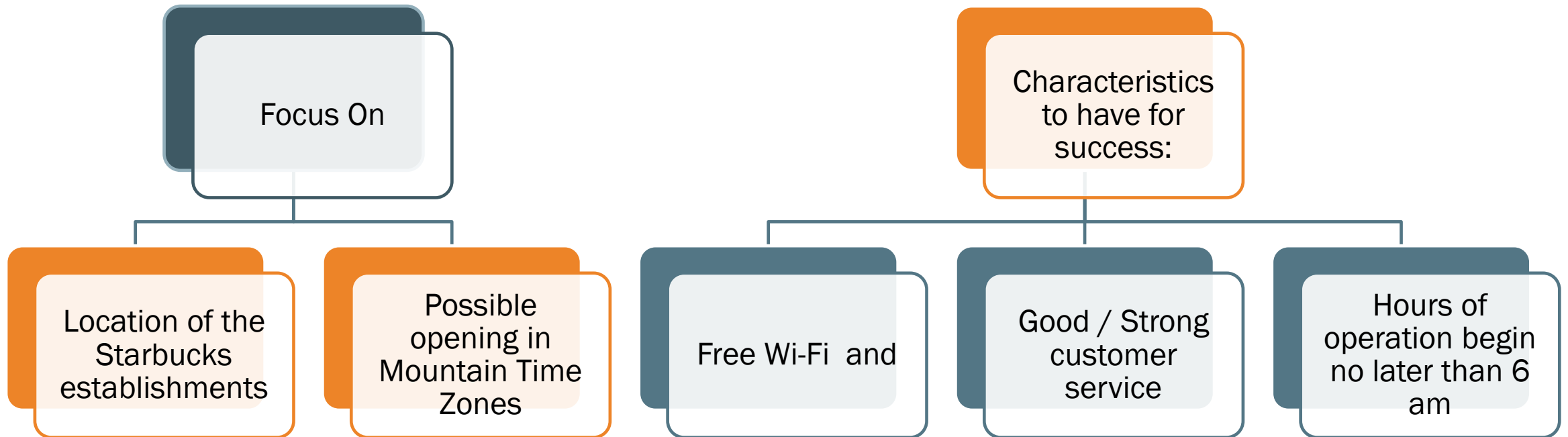
ADDITIONAL METRICS

- Investigate:
  - How the presence of a Starbucks affects the business status of other coffee shops in the area?

NEXT STEPS: FOCUS / CHARACTERISTICS TO HAVE

Focus On

- Location of the Starbucks establishments
- Possible opening in Mountain Time Zones

Characteristics to have for success:

- Free Wi-Fi and
- Good / Strong customer service
- Hours of operation begin no later than 6 am

THANK YOU – COFFEE KING

THANK YOU