Business Analytics

Project: Restaurant Analysis

Submitted to:

Prof. Vikas Gadre

Date: 23rd November 2021

Submitted by: Group 1

Aastha (G011)

Dhanishtha Dhingra (G021)

Sanket Parida (G031)

Vineet Garga (G041)

Nishant Pal (G051)

Arekatla Harshit (G061)

Background Note (Industry Analysis)

The year of 2018-19 saw the food sector in the nation to provide employment to around seven million people. The organised food service industry, which accounts for only 35% of the entire market, paid a staggering Rs 18,000 crore in taxes in 2018-19. The number is anticipated to more than double if the unorganised sector becomes organised.

According to the NRAI IFSR 2019, the total market size of the Indian food service sector is expected to be Rs 4,23,865 crore in 2018-19, rising at a CAGR of 9% to Rs 5,99,782 crore by 2022-23.

In Bengaluru, the average monthly spends per household on eating out is Rs 3,586, which is more than the national average of Rs 2,500. The expansion of Bengaluru's food service business has been attributed to the city's cosmopolitan culture and the IT and ITES sector, which have combined to create a revolution in inhabitants' eating and cooking habits.

Introduction and Requirements Analysis

Bangalore is gradually establishing a reputation for being a high-paced metropolis that never stops moving. In keeping with this trend, a sizable portion of the city's working population increasingly orders food from restaurants and takeaway shops, both at home and at work.

While opening a new restaurant venture in Bangalore, following are some of the focus areas

1. Which location would allow the restaurant to attract a large number of people?
2. Which dishes and cuisines would attract the most customers?
3. What is the most appropriate price at which a particular meal should be offered?
4. What are the qualities that are most preferred by the customers and what are the qualities where they want changes?
5. How is the competition so that most appropriate market entry opportunity could be located?

This strategy would enable a restaurant owner to find the most appropriate opportunity to start a new restaurant.

Business Domain and Analysis

In this project, we performed a comprehensive analysis of the Zomato Bengaluru restaurant dataset from the perspective of opening a new restaurant. We evaluated the various prospects regarding restaurants such as cuisines, cost of meal etc., to find out the most feasible and appropriate options. We also performed sentiment analysis and topic modelling techniques on reviews to understand voice of customers.

Pre-Existing Issues & Inefficiencies

As many as 60% of new eateries fail during their early stages due to the severe competition. This includes but isn't limited to the restaurant's proximity to its target demographics and the sort of cuisine it serves. In order to ensure that a new restaurant is a success, all of these factors must be taken into account. Analysis of the restaurants in Bangalore will assist stakeholders make data-driven decisions on how best to open and operate their own restaurant. In order to set up a successful business model, we need to take into account a variety of factors, including the size of the market in the region, top performing locations, favourite cuisines, dishes loved, demand, the best pricing match, and most significantly, the feelings of the people. As a result, our goal here is to help the

stakeholder who is considering opening a new restaurant in Bengaluru by analyzing the data set and providing important insights.

Application of Analytics to overcome the challenges faced

To begin with, we focused on a descriptive study of prime locations, which were subsequently subdivided into restaurant kinds (e.g. steakhouses), further subdivided into cuisines, and ultimately an analysis of favourite meals. The second portion of the investigation was more qualitative in nature, with the goal of gauging people's feelings in order to pinpoint areas that need attention as well as those where progress may be made.

There were other issues that needed to be addressed as a stakeholder in the restaurant sector, such as determining what style of restaurant to build and where it should be located if the choice was made to proceed.

For the descriptive analysis aspect, the following questions were tried to be answered:

- Which would be the best area to open a new restaurant?
- What is the type of restaurant that should be opened?
- What are the dishes that one could offer in a restaurant that would help build its sales?
- Does one need to provide speciality or additional services to attract customers?
- Are they any other factors that are linked to the overall rating of the restaurant?

Now, for the qualitative analysis aspect, the following questions were tried to be answered:

- What insights from sentiment analysis were found to be positive?
- What aspects of the analysis were negative?
- What aspects can be improved from the analysis?

Implications of findings on stakeholders and the society

We were able to analyse a wide range of indicators thanks to the abundance of data in our project, including user ratings, online bookings, online deliveries, and so on. In this way, we were able to examine what makes a restaurant profitable, the sorts of restaurants available, and the varieties of cuisines served. And last but not least, the food that is being served. The stakeholder's perspective might supply him with aspects to improve upon, such as giving discounts to entice larger sales volume or operating a restaurant in a less competitive region. Using sentiment analysis, we can observe what customers think about the restaurant's strengths and disadvantages, which can help us make future decisions. For the society, it provides in-depth research of the sorts of food we eat and what we as human beings enjoy and dislike, which in turn may empower individuals to come up with new ideas.

Analysis method used with steps, scripts, screenshots, etc.

The tool that we decided to use was Alteryx Designer Suite. This enabled us to carry out the analysis on the dataset. To further conduct and in-depth analysis we used Alteryx Intelligence Suite which has a plethora of advanced analytics options.

The dataset was taken from Kaggle. Its link is as follows:

https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants

The original dataset was very muddled, haphazard and full of null elements, missing and incorrect data and other such issues. The first step was to clean the data and all the above-mentioned problems were removed and the data was ready for processing.

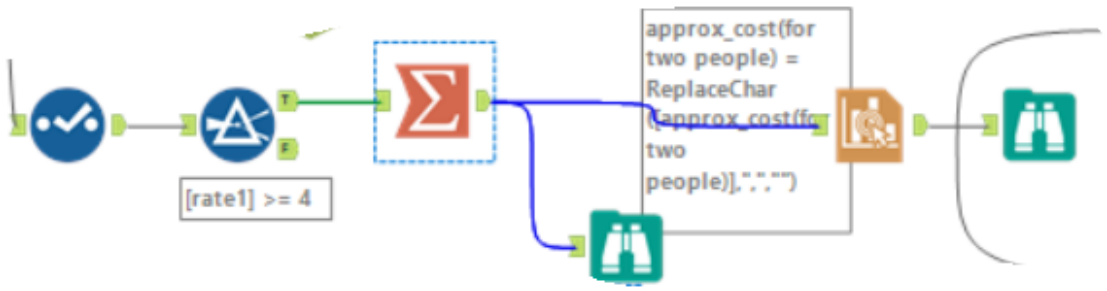
The data obtained from the above in the tabular column is as follows:

RecordID	url	address	name	online_or	book_tabl	rate	votes	phone	location	rest_type	dish_liked	cuisines	approx_c	reviews_li	menu_iter	listed_in(t	listed_in(c	rate1	rate2
1	https://wv	942, 21st M	Jalsa	Yes	Yes	4.1/5	775	080 42297	Banashank	Casual Din	Pasta, Lun	North Indi	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
1	https://wv	942, 21st M	Jalsa	Yes	Yes	4.1/5	775	080 42297	Banashank	Casual Din	Pasta, Lun	Mughlai	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
1	https://wv	942, 21st M	Jalsa	Yes	Yes	4.1/5	775	080 42297	Banashank	Casual Din	Pasta, Lun	Chinese	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
2	https://wv	2nd Floor,	Spice Elep	Yes	No	4.1/5	787	080 41714	Banashank	Casual Din	Momos, Lu	Chinese	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
2	https://wv	2nd Floor,	Spice Elep	Yes	No	4.1/5	787	080 41714	Banashank	Casual Din	Momos, Lu	North Indi	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
2	https://wv	2nd Floor,	Spice Elep	Yes	No	4.1/5	787	080 41714	Banashank	Casual Din	Momos, Lu	Thai	800	[('Rated 4.	[]	Buffet	Banashank	4.1	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Cafe	Churros, C	Cafe	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Casual Din	Churros, C	Cafe	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Cafe	Churros, C	Mexican	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Casual Din	Churros, C	Mexican	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Cafe	Churros, C	Italian	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
3	https://wv	1112, Next	San Churro	Yes	No	3.8/5	918	+91 96634	Banashank	Casual Din	Churros, C	Italian	800	[('Rated 3.	[]	Buffet	Banashank	3.8	5
4	https://wv	1st Floor, A	Addhuri Ud	No	No	3.7/5	88	+91 96200	Banashank	Quick Bite	Masala Do	South Indi	300	[('Rated 4.	[]	Buffet	Banashank	3.7	5
4	https://wv	1st Floor, A	Addhuri Ud	No	No	3.7/5	88	+91 96200	Banashank	Quick Bite	Masala Do	North Indi	300	[('Rated 4.	[]	Buffet	Banashank	3.7	5

Now, from the data, we didn't select a few features which we found not relevant to our analysis such as menu item, phone number, listed in (since it was a repeat of restaurant type).

After this, an ID called as record id has been created to help identify the record in the data set. Further, using the text to columns feature, various collated data such as restaurant type, rating (for e.g., 4.1/5), types of cuisines, have been split into multiple rows. Next, a data cleanser tool was applied to clean the data after doing all these operations. Along with this, certain columns such as approx. cost for 2 had comma values in them which made them store as a string instead of a floating number. Therefore, such anomalies were removed along with the new restaurants which didn't have any rating and hence, could not contribute to our analysis.

For our analysis, we decided to study the successful restaurants and for that, we took only restaurants that had a rating of equal to or greater than 4.



Now, using Summarizer, the data was grouped together as similarly done in a pivot table in excel. Then, and interactive chart tool was used to create graphs. The browse tool was used at every step to find out exactly what operations took place. The result obtained is as follows:

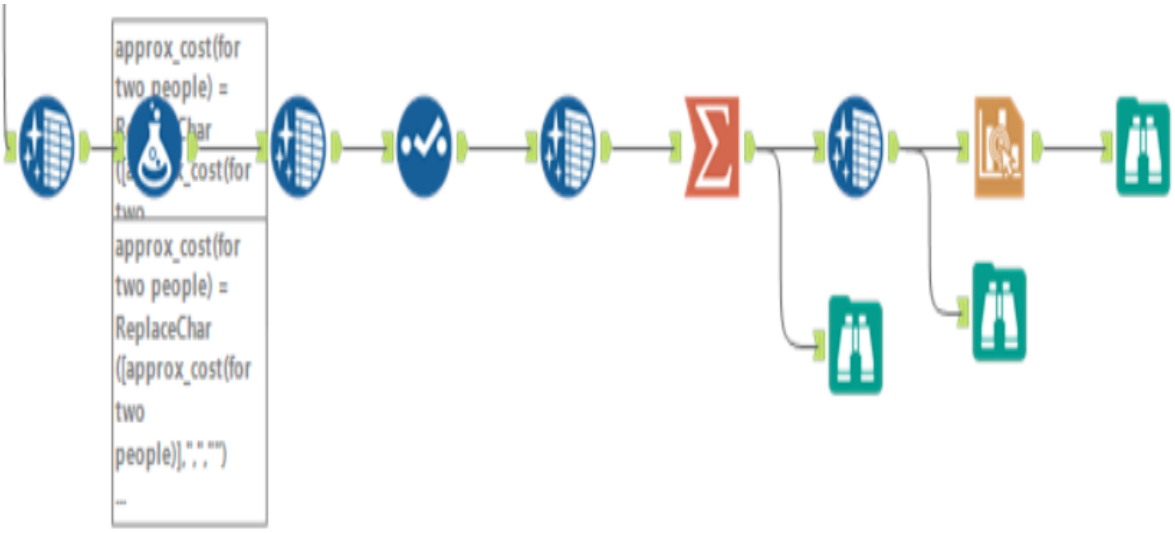
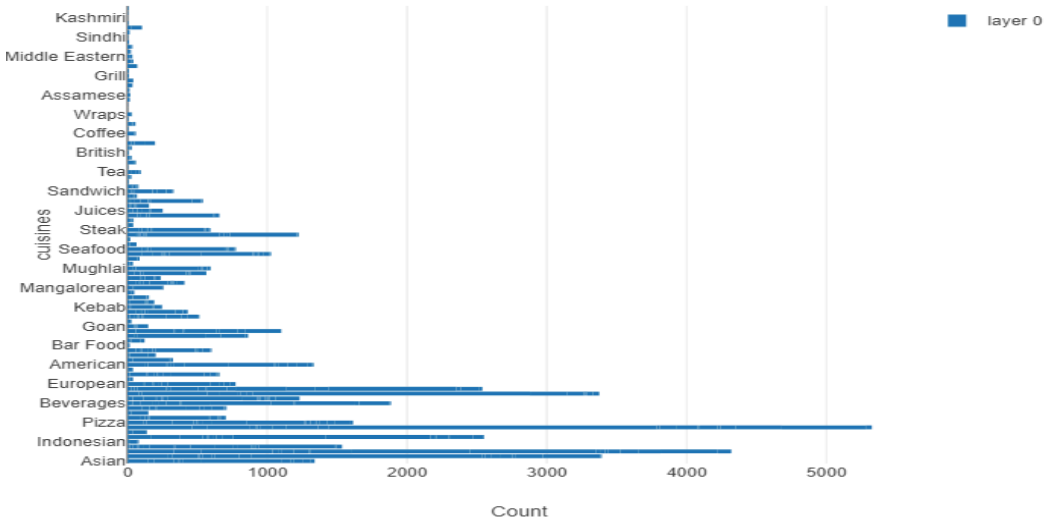
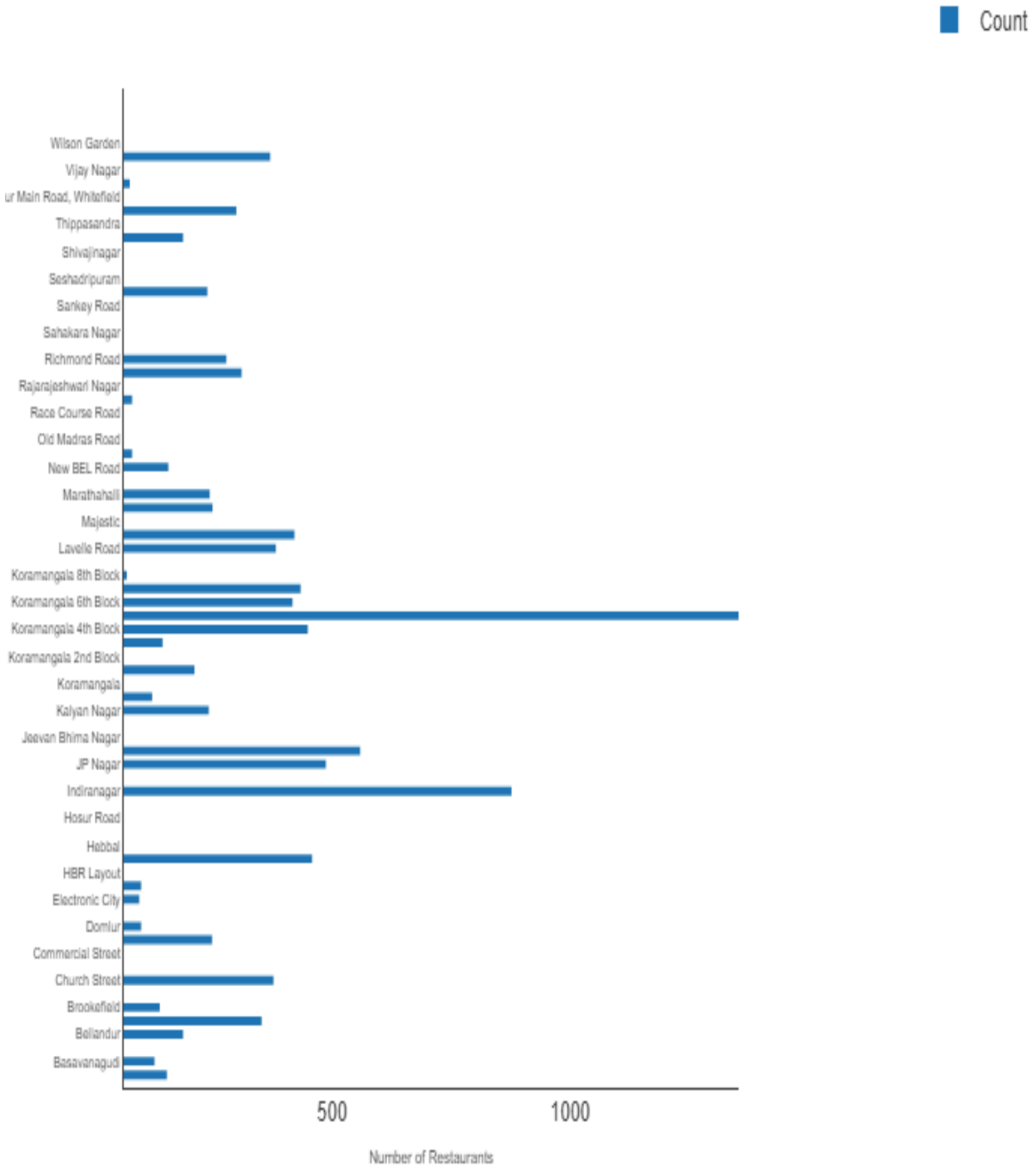


Figure showing the data cleaner tools along with the filters and selectors that have been added

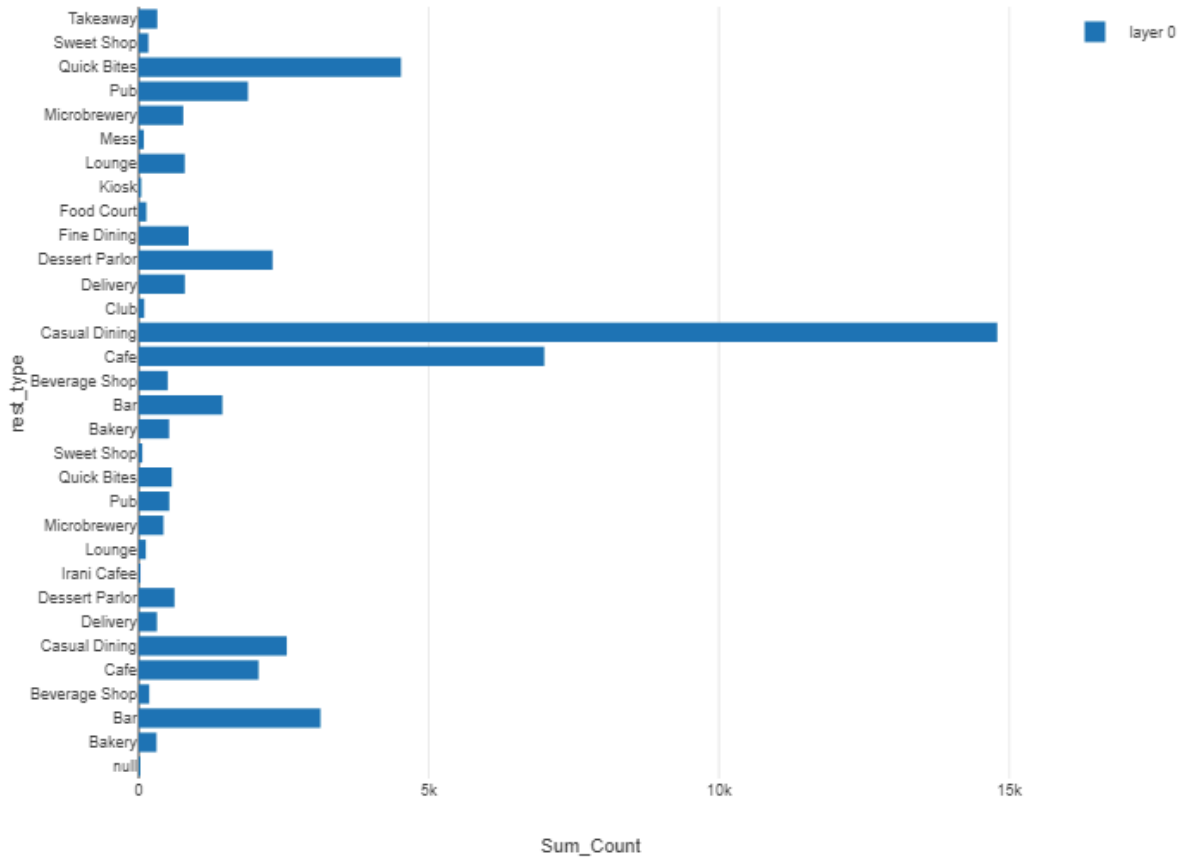
The interactive charts obtained are as follows:



The cuisine types that are highly successful shows **Continental, North Indian** as the leaders.



From the above, areas such a **Koramangala** and **Indiranagar** have the highest number of 4+ rated restaurants.



This figure shows the types of restaurants that are highly rated (4+), **Casual dining** leads the way.

Next to conduct and in-depth analysis, we took the leading restaurant types to highlight what kind of cuisines were most popular in those restaurants. It came out as:

Following are the top 4 restaurant types based on our analysis:

1. Casual Dining
2. Cafes
3. Quick Bites
4. Bar

We conducted in-depth analysis for each of the restaurant types in order to find out the preferred cuisine and dishes liked by the customers.

1. Casual Dining
   Cuisines liked:

We concluded that **North Indian** cuisine was the most preferred by the customers. So, we profiled the most liked dish in North Indian cuisine in a casual dining restaurant.

Dishes liked:

Similar process was carried out for the rest of the restaurant types and cuisines with the result as follows:

Cafe

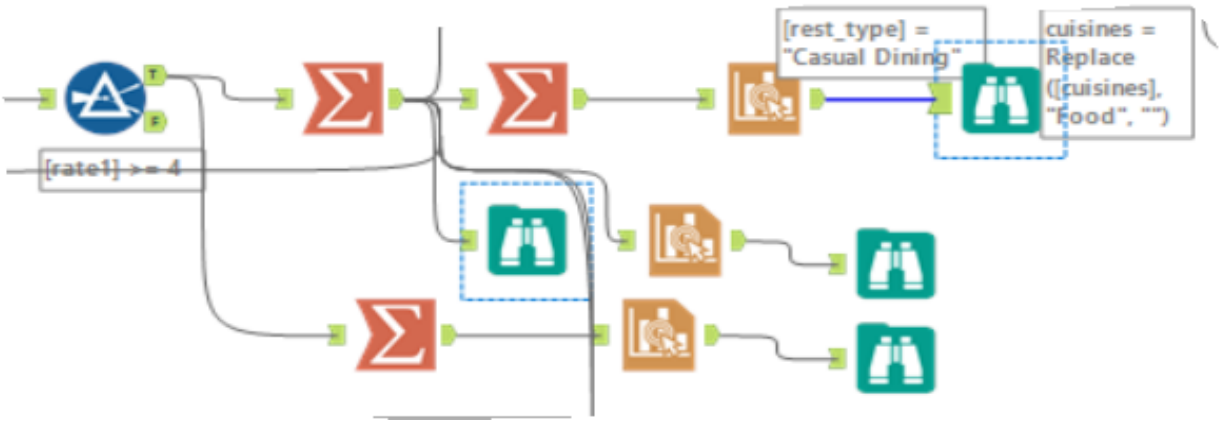
Continental

Quick Bites

Fast food

Bar **Chinese**

Word cloud figures showcasing the popular cuisines as per restaurant types of Casual Dining, Café, Quick Bites and Bar respectively.



Next, we decided to calculate the factors that are affecting the rating. It was decided that the rating of a restaurant was as close as to a NPS score and hence, it was the main factor that was contributing to a restaurant's success. Now to find out what factors are to be taken, we used number of votes, online delivery, table booking and approximate cost for two as our inputs. The online ordering and table booking parameters were converted to a binary data type, since it only had two responses.

online_order	book_table	votes	location	approx_cost(for two people)	rate1
1	1	775	Banashankari	800	4.1
1	1	775	Banashankari	800	4.1
1	1	775	Banashankari	800	4.1
1	0	787	Banashankari	800	4.1
1	0	787	Banashankari	800	4.1
1	0	787	Banashankari	800	4.1
1	0	918	Banashankari	800	3.8
1	0	918	Banashankari	800	3.8
1	0	918	Banashankari	800	3.8
0	0	88	Banashankari	300	3.7
0	0	88	Banashankari	300	3.7
0	0	166	Basavanagudi	600	3.8
0	0	166	Basavanagudi	600	3.8
1	0	286	Basavanagudi	600	3.8

Figure showcasing the dataset parameters used:

On conducting a linear regression on the same, we got a high intercept and a high value of residual signifying that the solution wasn't optimal. Next, we decided to conduct a decision tree regression. 20% of dataset was taken as sample data and the decision tree regression model so calculated was then tested on the 80% of data to get the actual vs predicted parameters of our model.

Using the model comparision tool, we got the parameter Root Mean Square Error as **0.3448**

Variable	Importance	Impact
votes	58.926%	
approx_cost.for.two.people.	24.188%	
book_table	13.705%	
online_order	3.181%	

Figure showcasing the importance of different parameters contributing to overall rating of a restaurant

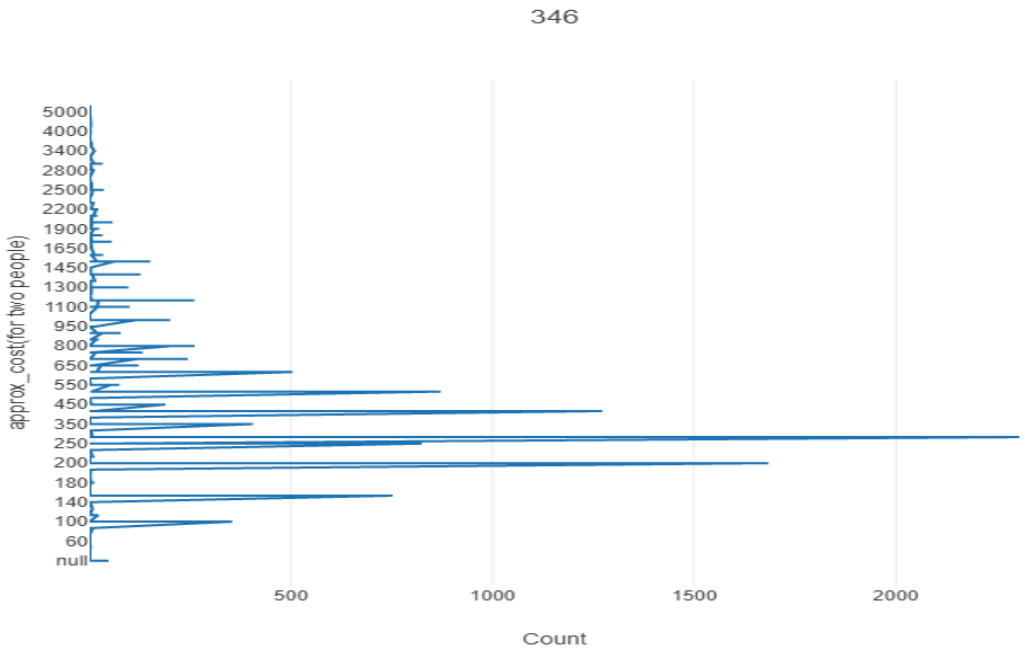
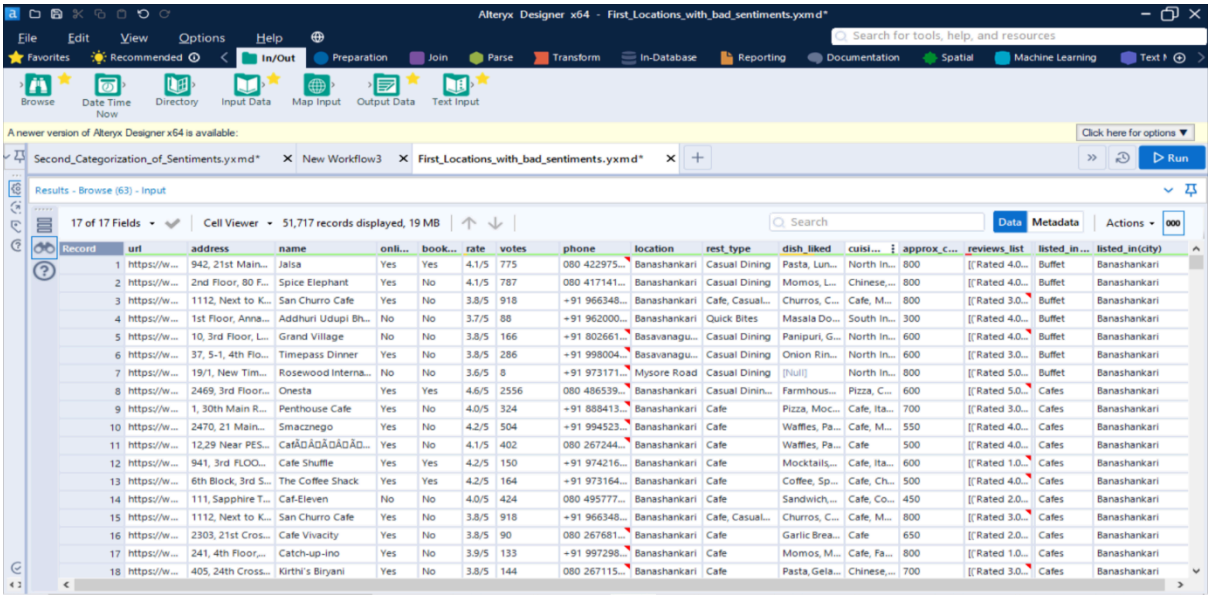


Figure showcasing the approximate cost for two vs number of restaurants rated above 4.

Sentiment Analysis

- **Dataset**: As shown below the dataset consists of URL of the restaurant on Zomato, address, name, online order, book table, phone no, location, restaurant type, approximate cost for two, reviews list, the dishes liked as columns for each row.



- **Data Cleaning:** We cleaned the data by converting the columns to the appropriate types such as online order and book table to Boolean types. Many unwanted characters were cleaned from the reviews list using regex.

- **Stemming and removal of stop words**: Stop words such as a, an, the etc., were removed and the remaining words were converted to root words.

- **Sentiment Analysis:** Using VADER((**Valence Aware Dictionary for Sentiment Reasoning**) algorithm, sentiments for each review were calculated through Sentiment analysis tool of Alteryx.

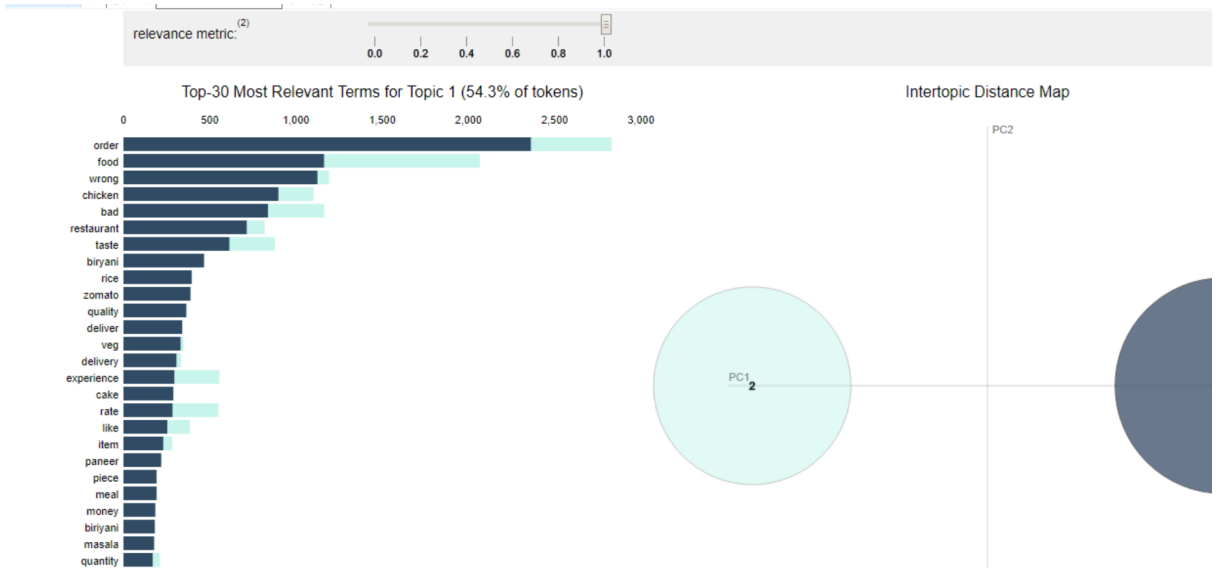
- **Segregation of negative reviews:** We filtered out the negative reviews.

13 of 13 Fields | 4,811 records displayed, 742 KB | Search | Data | Metadata | Actions

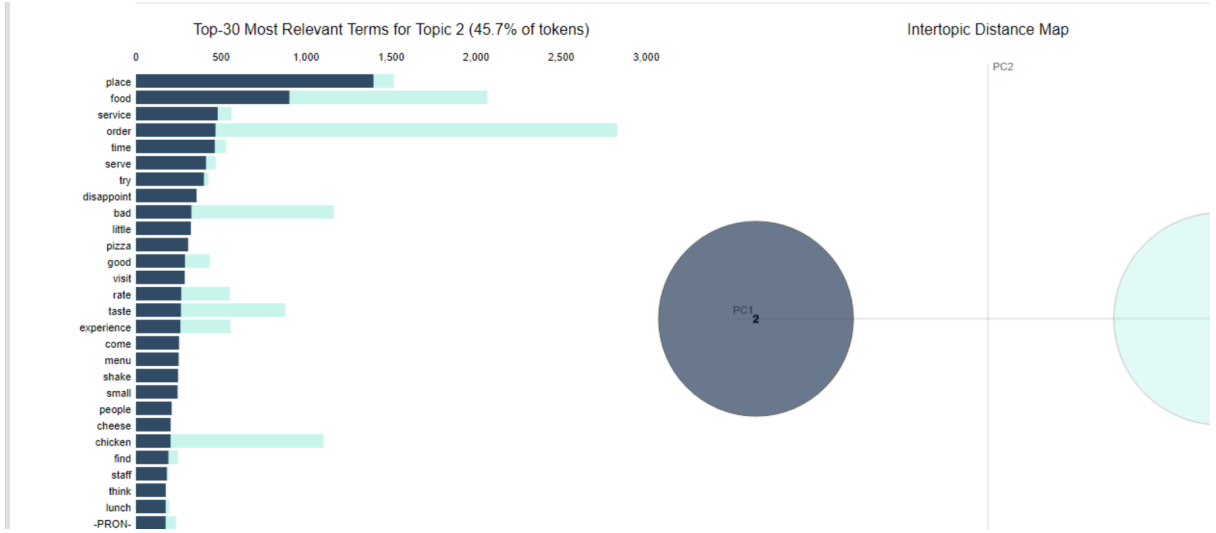
Record	e	listed_in_city	reviews_list	Review_sentiment_score	sentiment_category
1	es	Banashankari	horrible kid order bbq chicken wing chef special...	-0.2023	negative
2	es	Banashankari	hookah cafe mention page hookah place s separ...	-0.5423	negative
3	es	Banashankari	quality coffee wrong dilu	-0.4767	negative
4	es	Banashankari	tasty idlis kesari bath chatni sambar upto mark es...	-0.5829	negative
5	very	Banashankari	wrong restaurant veg manchurian worth fill tiny c...	-0.6486	negative

- **Topic modelling**: For these locations, we consolidated the negative reviews and applied topic modelling tool using LDA algorithm. We found out two clusters.

<u>Cluster 1</u>



<u>Cluster 2</u>



- **Analysis:**
  - Most frequent words in the first cluster were Order, Chicken, Place, Biryani, Rice, Try, Veg, Fry, Bad, Pizza, Shake, paneer.
  - Most frequent words in the second cluster were Food, Order, Wrong, Bad, Restaurant, Taste, Rate, Place, Service, Zomato, Quality, delivery.

- **Result:** We inferred from this is that the first cluster represented dishes with which the customer was unsatisfied these were Chicken, Biryani, Rice, Veg fry, Pizza, Shake, Paneer.

  The second cluster was about the attribute of service with which they were not happy with. These were Wrong order, taste, rate, Restaurant, Zomato, Quality, delivery.

Final Insights and Recommendations

After we have obtained all these insights, the following can be answered now:

Q1. What should be the location of a new restaurant?

- As per our analysis, Koramangala and Indiranagar have the cluster of restaurants that are highly rated. Either the stakeholder now may enter into this area or in a different untapped location depending on his business goals.

Q2. How should I price my restaurant in Bengaluru?

- The prices that show maximum positive rating correlation is approximate cost of two for **250.** Hence, the price should be kept around this range to improve the success rates.

Q3. What and which restaurant type, cuisine type and dishes should I offer?

- As per the word clouds, casual dining, café, quick bites and bar kind of restaurants have the most success. Next, the top cuisines on offer include North Indian, Continental, Fast food and Chinese. After further bifurcation, the top dishes are related to chicken, pizza, pasta, etc depending on the antecedent factors.

Q4. How can I improve the rating of my restaurant?

- A rating is the closest indicator of a restaurant's success as per the data set. The votes obtained have the maximum influence on the rating of eh restaurant. As such, it can be recommended that the restaurant converts its customers into giving votes and writing reviews for them, for example, they can offer a discount or free beverage if the customer decides to give them an honest rating. This is followed by the pricing strategy of the restaurant, and finally the table booking facility and online ordering parameter.

Q5. What are the dishes which I can improve so that I can get better customer satisfaction and be known for the quality?

- As from analysis the dishes which the customers need improvement are Chicken, Biryani, Rice, Veg fry, Pizza, Shake, Paneer. These dishes provide an opportunity for the restaurant to pioneer in because the current alternatives are not providing the right quality.

Q6. What are the parameters which the customer is not satisfied with the current alternatives present?

- From our analysis we gathered that order correctness, good taste, better rate, quality, and efficient delivery were the main parameters that the customer was concerned with. If the

new restaurant pioneers these parameters. It would become the most preferred restaurant in that place.

Exhibits

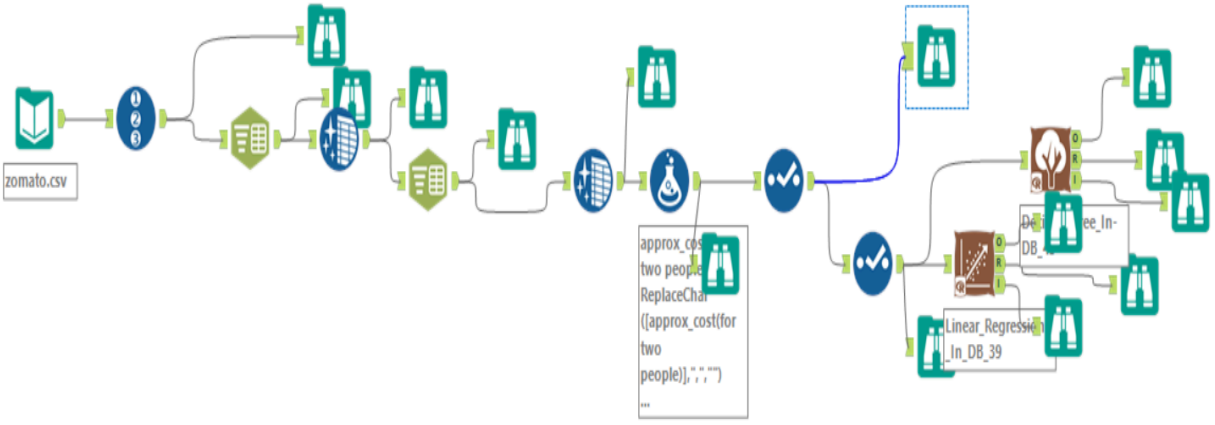


Figure for regression analysis

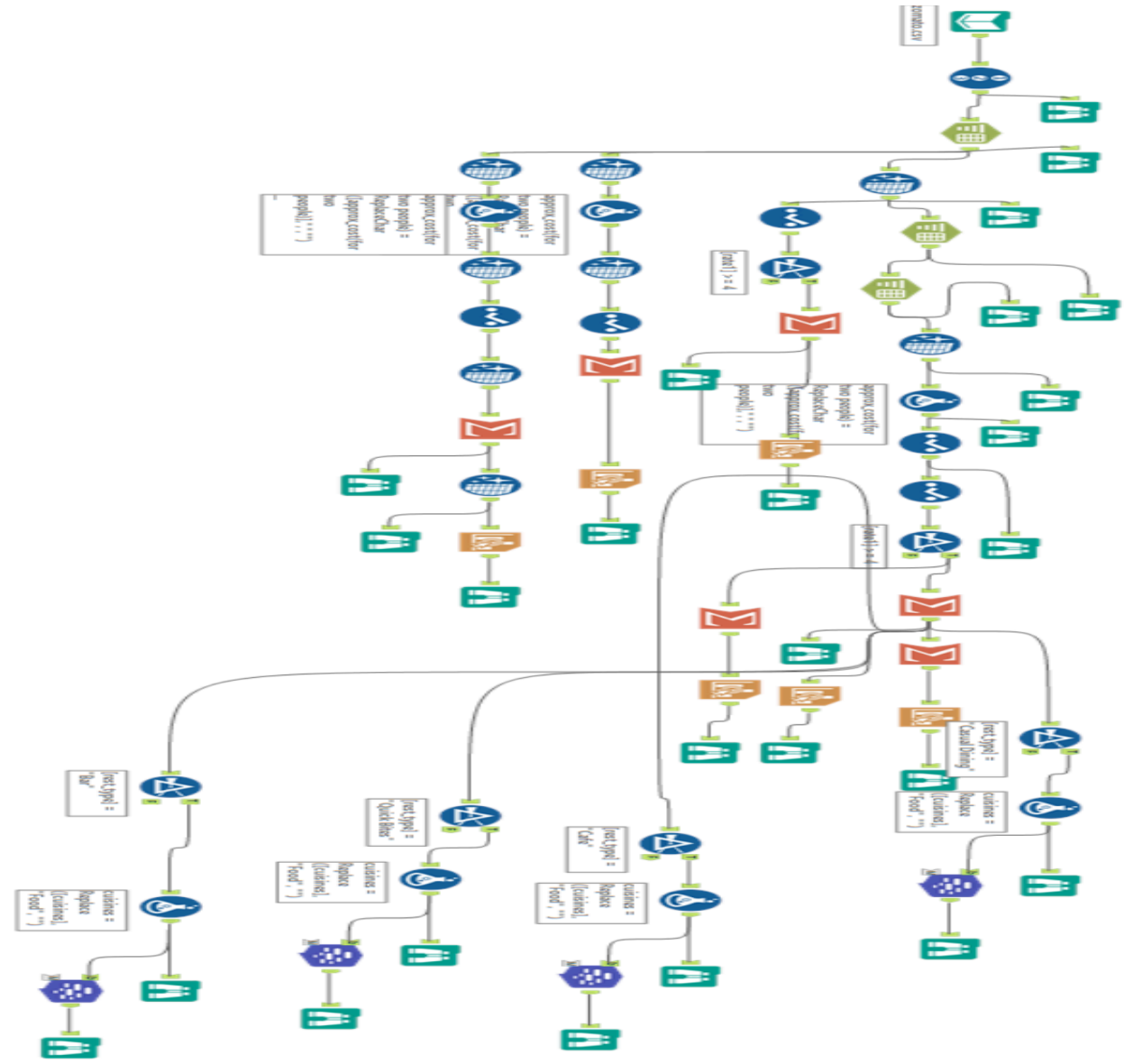
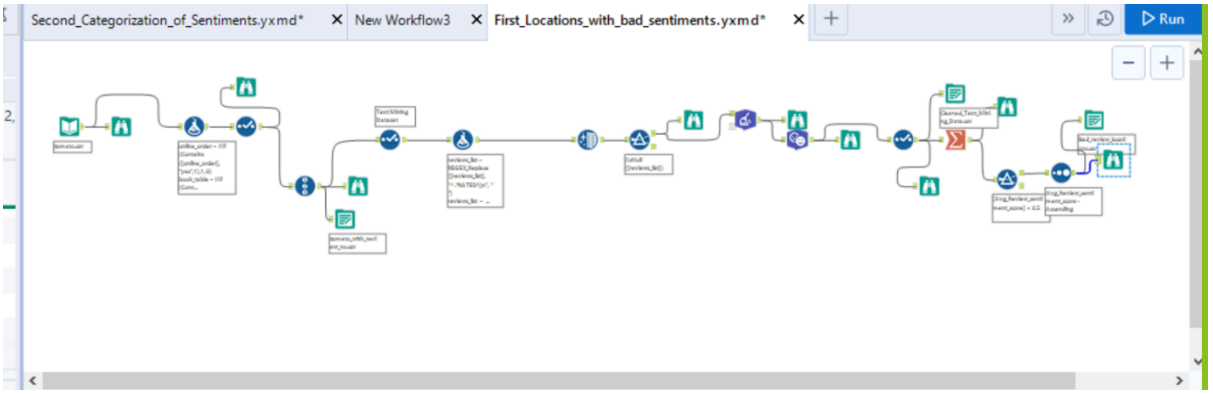
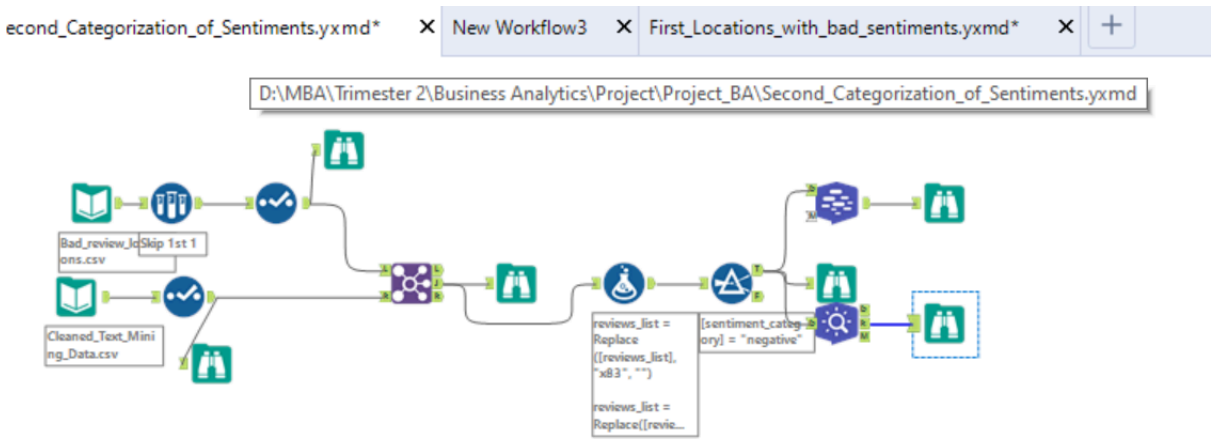


Figure for Text Mining, Word Cloud and Rating Analysis



Workflow of Sentiment Analysis



Workflow of Topic Modelling

References

- Why And How to Use Data to Run your Business More Effectively | Retrieved 25th October 2021 https://dgenious.com/restaurant-analytics/
- Issues and Challenges by Restaurant Industry | Retrieved 25th October 2021 https://www.restaurantindia.in/article/Issues-and-Challenges-by-Restaurant-Industry.6121
- Predictive Analytics using Alteryx by Lovelytics | Retrieved November 3rd 2021 https://www.youtube.com/watch?v=ia7MDt54JJc&t=2304s&ab_channel=Lovelytics
- Dataset – Zomato Restaurants Bangalore – Kaggle | Retrieved November 10th 2021 https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants
- Industry Analysis | Retrieved on 10th November 2021 | https://www.indiaretailing.com/2019/05/22/food/food-service/bengaluru-organised-food-service-market-3rd-highest-in-indian-metros-at-rs-20014-cr-nrai-report/
- Industry Analysis | Retrieved 16th November 2021 | https://nrai.org/bengalurus-organised-food-service-market-estimated-at-rs-20014-crore/
- Alteryx learning recourse | Retrieved 3rd November 2021 | https://www.youtube.com/watch?v=YISWJ-G7Zt4&t=1s&ab_channel=SimonSezIT