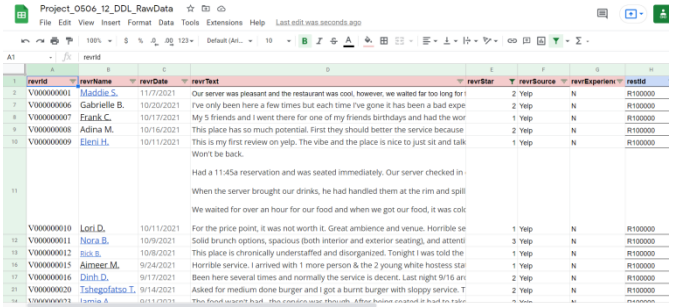
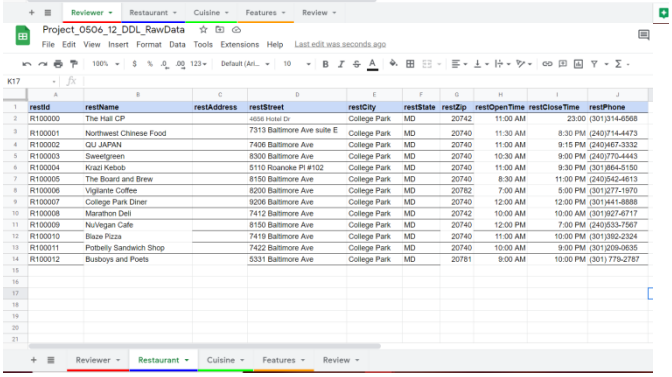
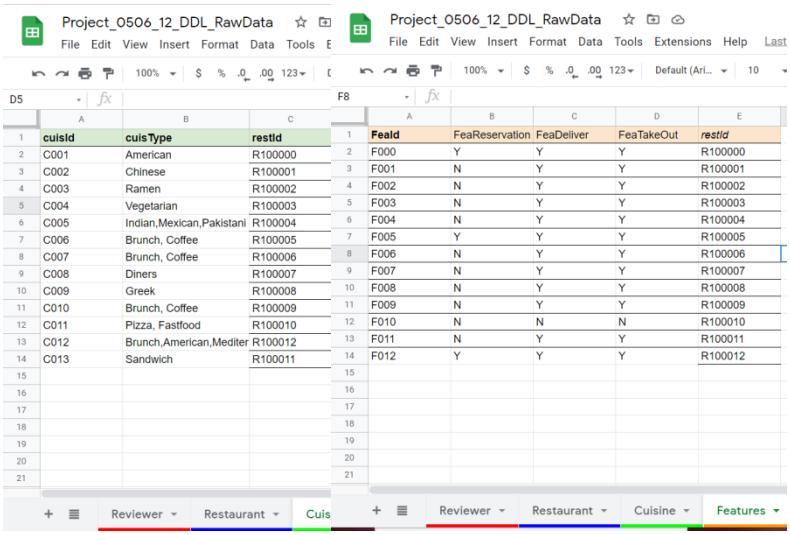
Background

- Our data is centralized on our users, who are individuals who submitted reviews on Yelp, Google, and Tripadvisor for our selected College Park, MD restaurants in the past year. With these users data in our database, our main goal is to provide customers seeking to dine at a restaurant in College Park, MD the insight and data they require in making an optimal decision
    - Example websites
    - https://www.yelp.com/biz/potomac-pizza-college-park
    - https://www.google.com/search?q=googel+reviews+potomac+pizza&rlz=1C1CHBF_enUS762US762&oq=googel+reviews+potomac+pizza&aqs=chrome..69i57j33i22i29i30.3870j0j4&sourceid=chrome&ie=UTF-8
    - https://www.tripadvisor.com/Restaurant_Review-g41078-d12866999-Reviews-Potomac_Pizza-College_Park_Maryland.html
- Our data was collected from these review-based websites and transported into a Microsoft Excel file. In this file, we sorted and labeled our input data into different sheets, one sheet per entity, and then transferred this file into our SQL database. This input data is broken down into 4 entities: Restaurant, Reviewer, Cuisine, and Features. Our group believes that these four entities and their relationships to one another provide the optimal tables and format for our SQL database to be of most use to our end-users.
-



-



- 

Mission statements:

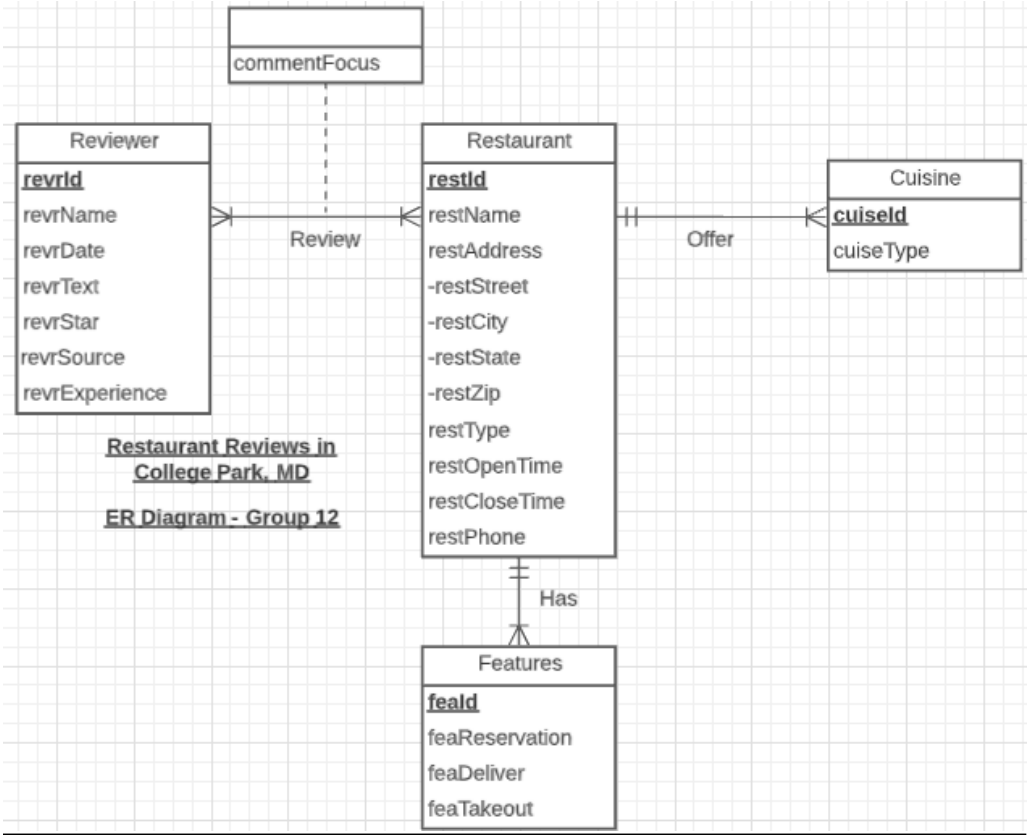
- Our mission statements were created to be simple yet descriptive enough to complete them in our final deliverable and be able to pull tangible results. Our main goal throughout this project, reflective in both mission statements, is to collect customer experiences via online reviews, and offer future individuals in the College Park, MD area seeking to dine a set of reviews and restaurant data that they can filter to their liking as they seek their ideal restaurant. Our mission statements each define the task at hand, are brief, and are defined by the scope of our selected restaurants and reviews in College Park, MD.
- 1) To analyze online reviews of restaurants in College Park, MD in order to gain insights on restaurant ratings and customers
- 2) To collect, sort, and analyze the comment focuses of online restaurant reviews in order to assist database users in ordering from College Park, MD owned restaurants based on different dimensions

Mission objectives:

- Our mission objectives were designed with our end-users in mind. With each of our objectives, we seek to answer key user questions and provide some of the most important results that our database can provide users seeking to find their optimal College Park, MD restaurant. Our mission objectives are each supported by our database and the data alike as well as converted to business transactions, which are represented below. Our project and database offer our users a multitude of ways to filter and define the selected College Park, MD restaurants. Along with our core mission objectives, which we developed for our users, our database offers our users restaurant-review focused results as well as delivering on answering our mission statements
- To find the positive feedback rate of each restaurant in 2021 (4-5 Stars)
- To find the restaurant that offers the most features in College Park, MD
- To find the average rating of each restaurant
- To find which aspects of each restaurant are rated positively and negatively

- Conceptual Database Design:

We have categorized our databases into different sections as per the ER Diagram designed above. The data has been split into 4 entities named Restaurant, Reviewer, Features and Cuisine.
There are 3 relations: One called Review which is a Many to Many relation between Reviewer and Restaurant. One is Offer which is a One to Many relation between Restaurant and Cuisine and the last one is Has which is a One to Many relation between Restaurant and Features.



- Logical Database Design

Relations:

Restaurant (**restId**, restName, restStreet, restCity, restState, restZip, restOpenTime, restCloseTime, restPhone)

Reviewer (**revrId**, revrName, revrDate, revrText, revrStar, revrSource, revrExperience, *restId*)

Cuisine (**cuisId**, cuisType, *restId*)

Features(**feaId**, feaReservation, feaDeliver, feaTakeOut, *restId*)

Review (***restId***, ***revrId***, commentFocus)

- Physical Database Design

Once we create the logical physical database design, we could create physical database design later on. Physical database is to store the data and ensure data integrity, security and recoverability.

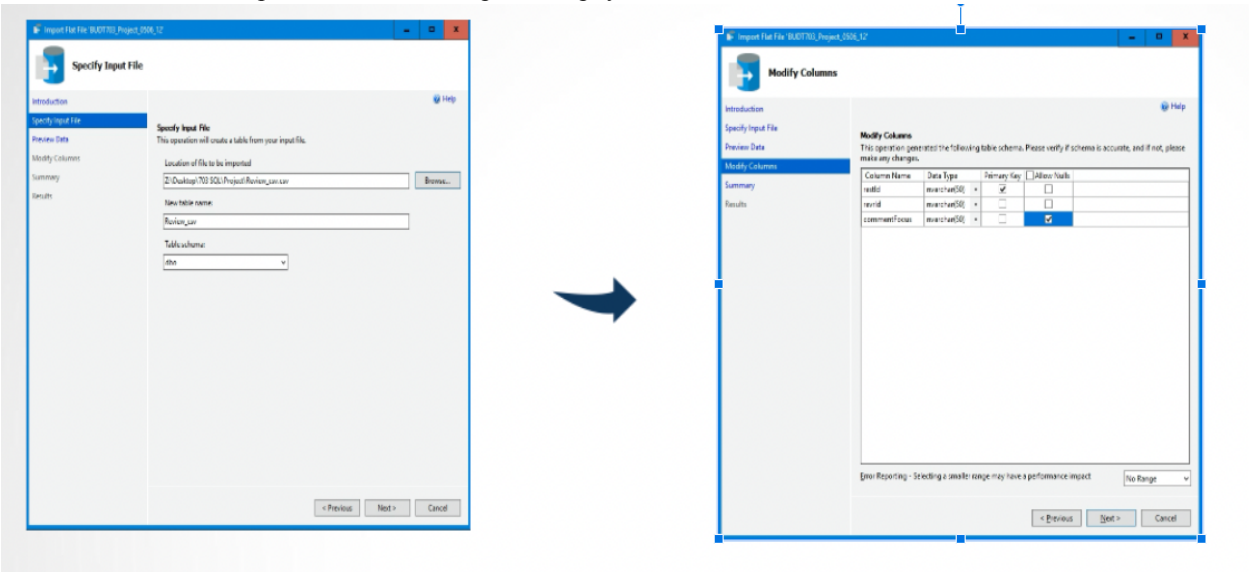
1. Create a table directly. We create the table using SQL database for restaurant, features and cuisine, adding appropriate data type and then add the foreign key at the end.

Create Table

```
CREATE TABLE [G12.Cuisine] (
      cuisId CHAR(4),
      cuisType VARCHAR(20),
      restId CHAR(10),
      CONSTRAINT pk_Cuisine_cuisId PRIMARY KEY (cuisId),
      CONSTRAINT fk_Cuisine_restId FOREIGN KEY (restId)
            REFERENCES [G12.Restaurant](restId)
            ON DELETE CASCADE ON UPDATE CASCADE)
```

	cuisId	cuisType	restId
1	C001	American	R100000
2	C002	Chinese	R100001
3	C003	Ramen	R100002
4	C004	Vegetarian	R100003
5	C005	Indian & Mexican	R100004
6	C006	Brunch	R100005
7	C007	Brunch	R100006
8	C008	Diners	R100007
9	C009	Greek	R100008
10	C010	Brunch	R100009
11	C011	Fastfood	R100010
12	C012	Brunch	R100012
13	C013	Sandwich	R100011

2. Import CSV file to complete the physical database.



We could set the primary key and set it not null for the primary key. However, we could not change the data type in this process and add the foreign key, so we chose to add the ALTER in SQL query.

```
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrId CHAR(10) NOT NULL;
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrName VARCHAR(50);
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrDate DATE;
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrText VARCHAR(8000);
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrStar integer;
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrSource VARCHAR(20);
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN revrExperience_Positive_or_Negative CHAR(2);
ALTER TABLE [G12.Reviewer]
 ALTER COLUMN restId CHAR(10);

ALTER TABLE [G12.Reviewer] ADD FOREIGN KEY (restId) REFERENCES [G12.restaurant](restId)
```

From now, we accomplish all the database preparation, including the logical database and physical database.

- Business Transactions

  Generally, four business transactions are applied to our project to further achieve the mission objective, in the way that analyzes the restaurants from different dimensions. This section will be carried out in terms of why this dimension was chosen to analyze the restaurant and how the analysis was developed and tested.

1) What is the average star rate of each restaurant?

One of the reasons for the design of the average star rate is that it can take advantage of the characteristics of all review scores and can roughly assess the general level of the restaurant. Moreover, considering the fact that all of the review scores(Star) are within a narrow range, from 1 to 5, so that the mean value is less likely influenced by outliers. In addition, in mathematics, the mean value could minimize the sum of squared errors significantly. Therefore, finding the average score of each restaurant is relatively objective and valid.

As shown in the code on the right, to find the average star of each restaurant, we need to first aggregate the data by restaurant name (restName), and then find the restaurant name by linking the two relations 'restaurant' and 'reviewer' relations.

```
--Q1: What the average star rate of each restaurant
SELECT t.restName,AVG(r.revrStar) 'avgRevrStar'
FROM [G12.Reviewer] r, [G12.Restaurant] t
WHERE r.restId =t.restId
GROUP BY t.restName
```

2) How many negative/positive reviews each restaurant received? And what about their rate?

The positive review rate is a direct indicator to infer the quality of a restaurant. Psychologically, the higher the positive review rate, the easier consumers to consume, and in turn, the easier to leave positive feedback. This is what marketing always emphasized in terms of a virtuous circle on customer experience .

As shown in the code on the right, "WHERE revrExperience_Positive_or Negative = 'P" and ' HAVING revrExperience_Positive_or Negative = 'N' ' is applied to count the number of negative and positive comments. Since the count( ) function is applied, the data needs to be aggregated. The results are shown in the most inner and secondary layer subquery respectively.
Then repeat twice 'ROUND(CAST (expression AS [data type]))/(expression),2)' to calculate the percentage with the number of bad reviews and the number of good reviews as the numerator respectively

```
GO
CREATE VIEW V2
AS
SELECT t1.restName, q1.cn 'Total reviews', q1.neg 'Number of Negative',ROUND(CAST
  (q1.neg AS FLOAT)/q1.cn,2) 'Negative Rate', q1.pos 'Number of Positive', ROUND
  (CAST(q1.pos AS FLOAT)/q1.cn ,2) 'Positive Rate'
FROM [G12.Restaurant] t1,(
    SELECT r3.restId, COUNT(r3.revrId) 'cn', q.ne 'neg',q.po 'pos'
    FROM [G12.Reviewer] r3, (
        SELECT r2.restId, COUNT(revrExperience_Positive_or_Negative ) ne, p.po
        FROM [G12.Reviewer] r2, (
            SELECT r1.restId, COUNT(revrExperience_Positive_or_Negative ) po
            FROM [G12.Reviewer] r1
            WHERE revrExperience_Positive_or_Negative = 'P'
            GROUP BY r1.restId ) p
        WHERE r2.restId = p.restId
        GROUP BY r2.restId, revrExperience_Positive_or_Negative , p.po
        HAVING revrExperience_Positive_or_Negative = 'N') q
    WHERE r3.restId = q.restId
    GROUP BY r3.restId, q.ne,q.po) q1
WHERE t1.restId = q1.restId
GROUP BY t1.restName, q1.cn,q1.neg,q1.pos
```

3）What is the distribution of reviews which are higher than 4 stars for each restaurant? The distribution is categorized within ‘environment/ food/ service’

This business transaction is mainly inspired by the 'Moments of Truth' theory(MOT)[1]. This theory was created by Jen Carlson, the former president of Scandinavian Airlines.MOT is the moment when a customer/user interacts with a brand, product or service to form or change an impression about that particular brand, product or service. In other words, it is the moment when the customer comes into contact with various resources of the company. This moment determines the future success or failure of the company. Therefore, we designed a third dimension of analysis, that is, to analyze the key information of the reviews of customers with a satisfied attitude (rating higher than 4 stars)so that we can figure out which aspect of the restaurant leaves the most impression on them. As there are many key factors, in order to better find the advantage, we summarize the customer reviews according to the three aspects of 'food', 'service' and 'environment' that most consumers care about.

As shown in the code on the right, firstly, since each table is repeatedly filtered for comments with more than 4 points, we create a view (focus) to encapsulate this result for efficiency.

```
GO
CREATE VIEW focus
AS
SELECT y1.restId,v.commentFocus,Y1.revrId
FROM [G12.Review] v ,(
SELECT e.restId, e.revrId
FROM [G12.Reviewer] e
WHERE e.revrStar >3
GROUP BY e.restId, e.revrId) y1
WHERE y1.revrId =v.revrId
```

Then, As shown in the code on the right, apply HAVING f.commentFocus = 'FOOD'/'ENVIRONMENT'/'SERVICE' to filter the requirement and aggregate the data by GROUP BY statement, and then count the number of comments by COUNT( ) function. Finally, merge three tables by FULL OUTER JOIN statement.

```
SELECT g.restName, g.restId,ISNULL(sef.environment, 0) 'environment', ISNULL
  (sef.food,0) 'food', ISNULL(sef.service,0) 'service'
FROM [G12.Restaurant] g, (
    SELECT ISNULL(f1.restId, se.restId) 'restId', f1.food, se.environment,
      se.service
    FROM (
        SELECT i.restId , COUNT(f.commentFocus) 'food'
                FROM [G12.Restaurant] i, focus f
                WHERE I.restId =f.restId
                GROUP BY i.restId, f.commentFocus
                HAVING f.commentFocus ='FOOD') f1
        FULL JOIN (
            SELECT s.restId, s.service, e.environment
            FROM (
                SELECT i.restId , COUNT(f.commentFocus) 'service'
                FROM [G12.Restaurant] i, focus f
                WHERE I.restId =f.restId
                GROUP BY i.restId, f.commentFocus
                HAVING f.commentFocus ='SERVICE') s
            FULL OUTER JOIN (
                SELECT i.restId , COUNT(f.commentFocus) 'environment'
                FROM [G12.Restaurant] i, focus f
                WHERE I.restId =f.restId
                GROUP BY i.restId, f.commentFocus
                HAVING f.commentFocus ='ENVIRONMENT') e
            ON s.restId=e.restId) se
        ON se.restId = f1.restId) sef
WHERE g.restId=sef.restId
```

4) Which restaurant has the most complete features and where is the location?

Easy to understand, in this business transaction, we want to figure out the restaurant that provides reservation, takeout and delivery service. Restaurants with complete facilities are more likely to increase customer satisfaction.

As shown in the code on the right, apply WHERE f.feaReservation = 'Y' and f.feaDeliver = 'Y' and f.feaTakeOut = 'Y' to filter the restaurant that meets these 3 requirements and then to find the location by connecting 'restaurant' and a1 relation.

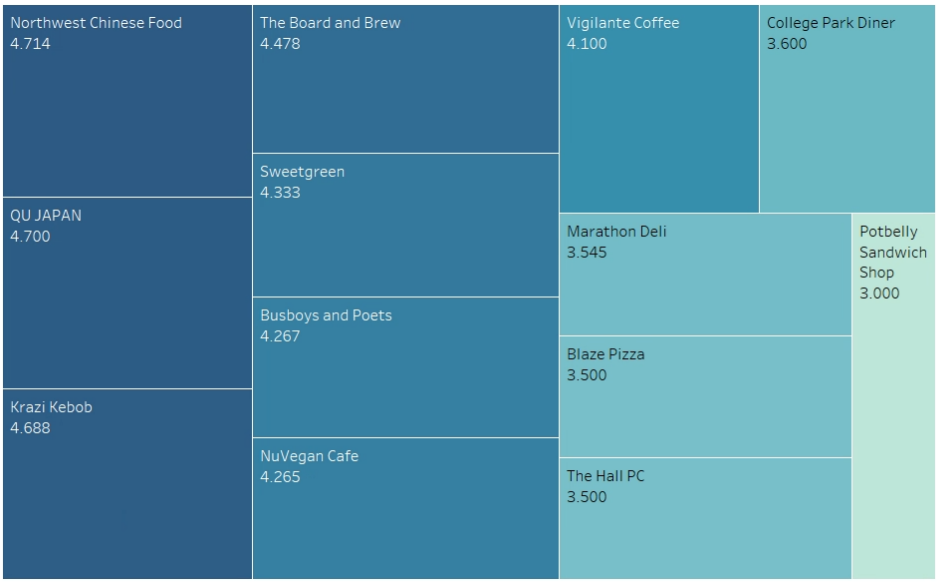
```
GO
CREATE VIEW V3
AS
SELECT t.restName, CONCAT(t.restStreet, ' ', t.restState) as 'Location'
FROM [G12.Restaurant] t,(
    SELECT f.restId
    FROM [G12.Features] f
    WHERE f.feaReservation = 'Y'
    and f.feaDeliver = 'Y'
    and f.feaTakeOut = 'Y') a1
WHERE t.restId = a1.restId
```

- Visualization

  Q1: What is the average star rate of each restaurant ?

Results | Messages

	restName	avgRevrStar
1	Blaze Pizza	3
2	Busboys and Poets	4
3	College Park Diner	3
4	Krazi Kebob	4
5	Marathon Deli	3
6	Northwest Chinese Food	4
7	NuVegan Cafe	4
8	Potbelly Sandwich Shop	3
9	QU JAPAN	4
10	Sweetgreen	4
11	The Board and Brew	4
12	The Hall PC	3
13	Vigilante Coffee	4

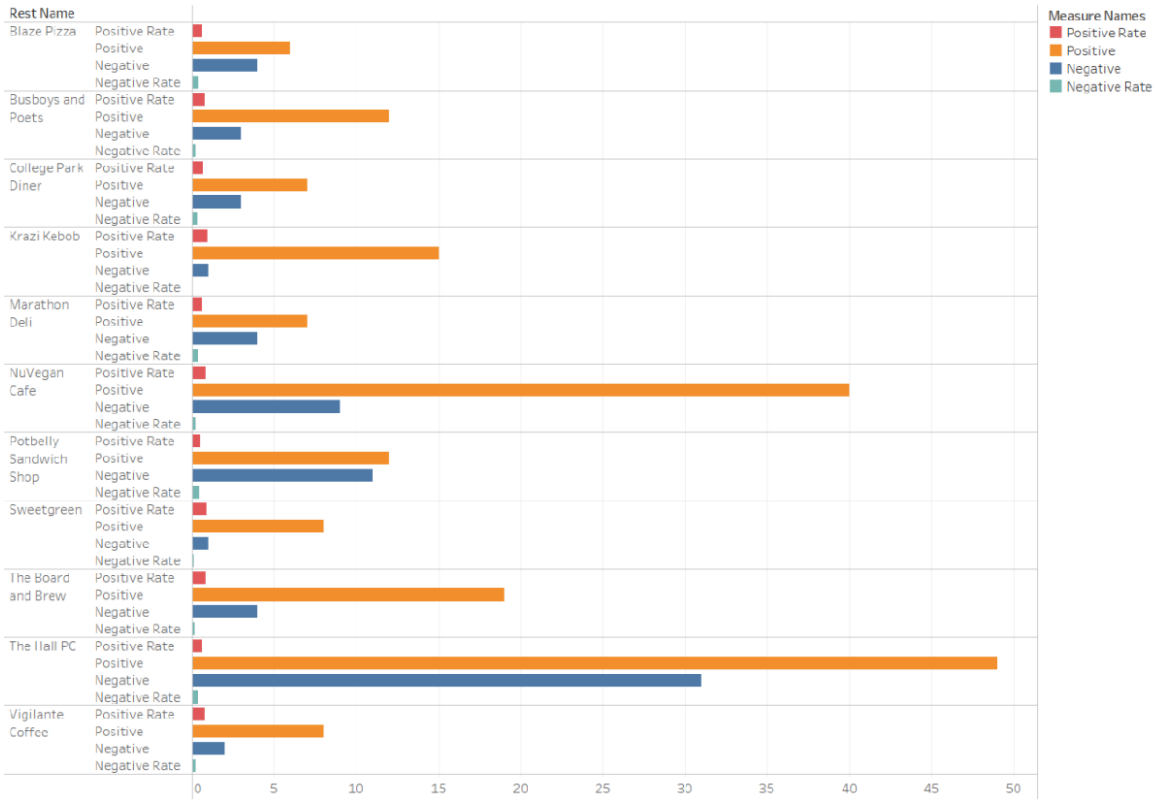


We use the treemap to visualize our first business transaction which presents the order of the average star by size and color. Concludingly, the average score of Northwest Chinese Food is higher than other restaurants,

2) How many negative/positive reviews each restaurant received? And what about their rate?

Results | Messages

	restName	Total reviews	Number of Negative	Negative Rate	Number of Positive	Positive Rate
1	Blaze Pizza	10	4	0.4	6	0.6
2	Busboys and Poets	15	3	0.2	12	0.8
3	College Park Diner	10	3	0.3	7	0.7
4	Krazi Kebob	16	1	0.06	15	0.94
5	Marathon Deli	11	4	0.36	7	0.64
6	NuVegan Cafe	49	9	0.18	40	0.82
7	Potbelly Sandwich Shop	23	11	0.48	12	0.52
8	Sweetgreen	9	1	0.11	8	0.89
9	The Board and Brew	23	4	0.17	19	0.83
10	The Hall PC	80	31	0.39	49	0.61
11	Vigilante Coffee	10	2	0.2	8	0.8



In order to visualize the number of negative/positive reviews in each restaurant received and their rate in Tableau, we should use calculation field and TYPE: IF [revrExperience Positive or Negative] = 'N' THEN 1 ELSE 0 END, to count the number of negative reviews;
TYPE: IF [revrExperience Positive or Negative] = 'P' THEN 1 ELSE 0 END, to count the number of positive reviews;
Then TYPE: SUM([Positive_Count]) / (SUM([Positive_Count])+SUM([Negative_Count])), to calculate the rate of positive reviews and negative reviews separately.
At the same time, we could generate the measure name for negative_count, positive_count, positive_rate and negative_rate. We drag them all into the “filter”, drag the negative_count, positive_count, positive_rate and negative_rate to columns one by one.
From the Tableau, Krazi Kebob (93.79%), Sweetgreen (88.89%) and The Board and Brew (82.61%) are the top 3 restaurants with the highest positive rate, while The Potbelly Sandwich Shop has the highest negative rate. Thus, we recommend our client reconsider whether to go to the Potbelly Sandwich Shop if they have this plan.

3) Which restaurant has the most complete features and where is the location?

Results | Messages

	restName	Location
1	Krazi Kebob	5110 Baltimore Ave MDCollege Park
2	Busboys and Poets	5331 Baltimore Ave MDCollege Park

Rest Name	
Busboys and Poets	5331 Baltimore AveCollege ParkMD20740
Krazi Kebob	5110 Baltimore AveCollege ParkMD20740

Since the location consists of street,city,state and zip, all of these factors should become a new measure in Tableau. More specifically, create a calculated field in the way: [Rest Street]+[Rest City]+[Rest State]+[Rest Zip] and rename the measure as 'Location'.

From the Tableau, Busboys and Poets, Krazi kebob are the restaurants that have completed features.

4) What is the distribution of reviews which are higher than 4 stars for each restaurant? The distribution is categorized within 'environment/ food/ service'

	restName	restId	environment	food	service
1	Blaze Pizza	R100010	0	5	1
2	Busboys and Poets	R100012	5	4	3
3	College Park Diner	R100007	0	4	0
4	Krazi Kebob	R100004	0	13	0
5	Marathon Deli	R100008	0	5	2
6	Northwest Chinese Food	R100001	0	7	0
7	NuVegan Cafe	R100009	7	31	2
8	Potbelly Sandwich Shop	R100011	0	7	3
9	QU JAPAN	R100002	3	6	1
10	Sweetgreen	R100003	0	7	1
11	The Board and Brew	R100005	6	7	6
12	The Hall PC	R100000	5	31	13
13	Vigilante Coffee	R100006	3	4	1

	Comment Focus		
Rest Name	ENVIRONMENT	FOOD	SERVICE
Blaze Pizza		5	1
Busboys and Poets	5	4	3
College Park Diner	3	4	
Krazi Kebob	2	13	
Marathon Deli		5	2
Northwest Chinese ..		7	
NuVegan Cafe	7	31	2
Potbelly Sandwich S..		7	3
QU JAPAN	3	6	1
Sweetgreen		7	1
The Board and Brew	6	7	6
The Hall PC	5	31	13
Vigilante Coffee	3	4	1

Highlight table is applied to show our results in Tableau, the darker the color, the more reviews mention that keyword, we could more likely make a hypothesis that the restaurant is highly possible good in this area. This table could give our client a reference on what impressed the reviewer most and further help them to make a decision by considering whether this highlight matches their own preference.For example, if our client is considering NuVegan Cafe and he is highly environmentally oriented person, we will recommend they reconsider that, since most of reviewers praise their food instead of the environment.

Basically and generally , this is how our database works.

Reference:

[1] Carlzon, Jan. *Moments of truth*. (1987) (BWB220141014 ed.). Cambridge, Mass.: Ballinger. Pub. Co.