Restaurant Data Analysis

Introduction:

For our SAS project, we will be analyzing 2012 Mexico restaurants data to provide insights for both restaurants and customers. Throughout this report, we will touch upon how we come up with our business problem after data inspection, discuss our approaches on data cleaning and validation using different means of measure, showcase visualizations, and conclude with findings and suggestions for future studies.

Data Description & Documentation:

The 2012 Mexico Restaurants Data consisted of data from restaurants and customer information/ratings. The restaurant data had accepted methods of payments, cuisines, hours open, parking options, and geographic information and other attributes. The data from the consumers consisted of cuisines of the restaurants, payment methods used, varied consumer information and lastly ratings of each restaurant based on service, food, and overall rating.

Business Problem:

After reviewing the data at our disposal, we formulated the following business problem:

> *"Customers in Mexico seek comprehensive, yet user-friendly recommendations for restaurants. They would like insights on ranks, ratings, and details on features of the restaurants that will help them make informed dining decisions."*

We will assist these customers by following these objectives:

1. Identify the top rated restaurants by assessing their service ratings, food quality ratings, and overall ratings.
2. Analyze different restaurant features (cuisines, price, dress code, etc.) for restaurants fulfilling certain rating conditions (top 10, average rating > 1.75 etc.).
3. Provide tailored recommendations for customers given their budget and preferences

Data Cleaning & Validation:

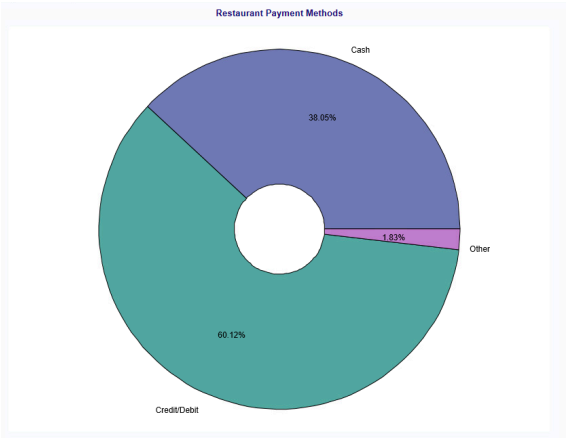
To get an overview of the variables and their observations requiring cleaning, we conducted a thorough inspection of each imported CSV file. One-way frequency reports were then created to assist in further investigation of unique values of certain variables. The following list outlines the steps taken in cleaning the data.

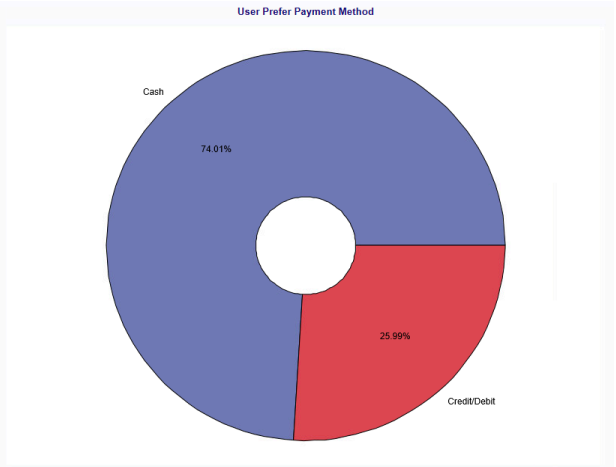
1. Common for all data sets with missing values were that they were recorded as question marks (?), so we re-coded these instances to nulls using empty strings since all variables with missing values were character variables.
2. In the data set geoplaces2, we imputed missing data in the *state* and *country* columns based on the information in the *city* column.
3. To accommodate for the inconsistencies in the way the city/state 'San Luis Potosi' was recorded, we re-coded all variations (slp, s.l.p., S.L.P, etc.) to spell out the full name of the *city*/*state* with proper capitalization.
4. We also cleaned other inconsistencies such as differences in capitalization for *country* using PROPCASE().
5. Similarly, we used UPCASE() for *franchise*, to provide a visually clearer differentiation between *t* and *f* (short for true and false).

6. Another measure we took to enhance readability of certain variables such as *alcohol*, was to replace underscores with spaces.
7. Specifically for the data set *geoplaces2* and *userprofile*, we dropped variables that were lacking context (e.g. *color*) or contained majorly or entirely missing values (e.g. *fax*).
8. For the data sets *userpayment* and *chefmozaccept*, we grouped the different payment methods into four categories (Cash, Check, Credit/Debit, and Gift Card) given that five original categories each made up less than 1% of the total observations. Additionally, we saw unique observations like 'VISA' and 'Visa', which were grouped into Credit/Debit for instance.
9. We created another way to consider the *ratings* data by recoding the values of 0,1, and 2 as 'Bad', 'Average', and 'Good' respectively.
10. For the *hours* data, we transposed the data from long to wide format. Initially, there were only three columns: placeID, hours, and days, with the placeID repeating for as many variations in hours that existed. By creating a column for each day of the week, the hours for a restaurant could be viewed on one single row.

Analysis & Visualizations:
Before creating visualizations, we first verified the cleaned data and reformatted variables when needed. Then, using our business problem as a guideline, we created different tables and graphs with graph tasks and proc sql. Based on the graphs and tables provided below, restaurants can find user preferences to improve their business while customers will be able to get a list of recommended restaurant names based on ratings and preferred characteristics.

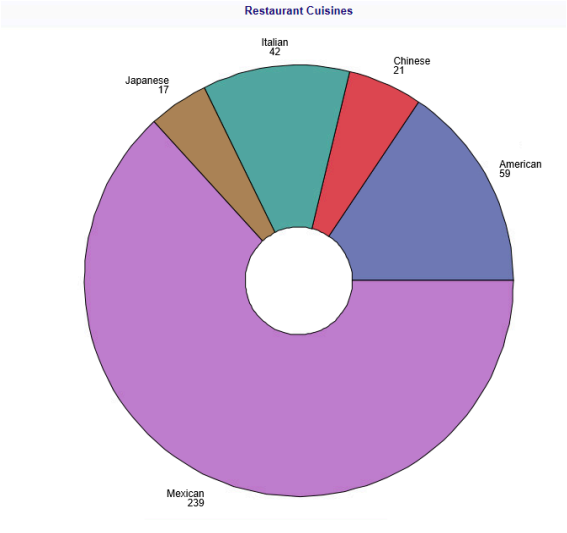
Preferred payment method: Two pie charts below show payment methods used by restaurants and users. The charts concluded that users and restaurants have different preferred payment methods.



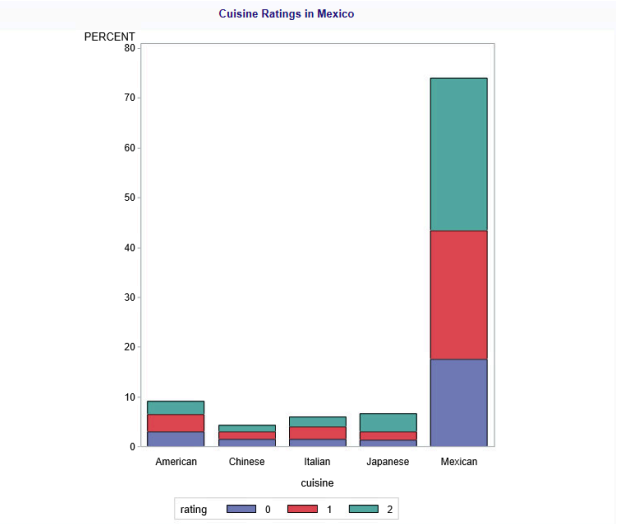


(Chart 1. Percentage of Payment Methods used) (Chart 2. Percentage of User Preferred Payment Methods)

Cuisine Ratings: Using summary statistics, we found the average rating for the most popular cuisine in Mexico, and created a pie chart and a bar graph as seen below. We found that the highest rated cuisine was Japanese while the most common was Mexican.



(Chart 3. Average Cuisines offered in Restaurants)



(Chart 4. Cuisine Ratings)

Rating by characteristic: A summary table was produced to represent restaurant characteristics and their average rating for each category.

Summary Tables
Ratings by Characteristics

	alcohol			dress_code			Rambience		price		
	Full Bar	No Alcohol Served	Wine-Beer	casual	formal	informal	familiar	quiet	high	low	medium
avg_rating	1.29	1.15	1.23	1.19	1.92	1.17	1.18	1.18	1.26	1.06	1.23
avg_food	1.22	1.19	1.26	1.26	1.75	1.20	1.22	1.13	1.25	1.14	1.26
avg_service	1.17	1.04	1.17	1.12	1.92	1.07	1.10	0.94	1.17	0.94	1.16

(Table 1. Average Ratings by Restaurant Characteristics)

Restaurant recommendations based on characteristics: the tables below were created using PROC SQL statements to return the names of the highest rated restaurants based on different customer preferences.

- Average rating over 1.75 for overall, food, and service

Best Restaurants on Average

name	avg_rating
Restaurant Las Mananitas	2
Michiko Restaurant Japones	2
Emilianos	2
Cafe Punta Del Cielo	1.833333
La Estrella De Dimas	1.8
Giovannis	1.75
Restaurante La Parroquia Potosina	1.75
Log Yin	1.75

Best Restaurants for Service

name	avg_service
Restaurant Las Mananitas	2
Cafe Punta Del Cielo	1.833333
Emilianos	1.8
El Cotorreo	1.75
Restaurante La Parroquia Potosina	1.75
Giovannis	1.75

Best Restaurants for Food

name	avg_food
Restaurant Las Mananitas	2
Giovannis	2
Little Pizza Emilio Portes Gil	2
Michiko Restaurant Japones	2
La Estrella De Dimas	2
Emilianos	1.8
Log Yin	1.75
Restaurante La Parroquia Potosina	1.75
El Oceano Dorado	1.75

(Table 2. List of best restaurants on average, for food, and service.)

- Based on alcohol with average rating over 1.5

Best Restaurants on Average with No Alcohol

name	avg_rating	alcohol
Michiko Restaurant Japones	2	No Alcohol Served
Cafe Punta Del Cielo	1.833333	No Alcohol Served
La Estrella De Dimas	1.8	No Alcohol Served
Giovannis	1.75	No Alcohol Served
Log Yin	1.75	No Alcohol Served
La Cochinita Pibil Restaurante Yucateco	1.714286	No Alcohol Served
Mariscos El Pescador	1.692308	No Alcohol Served
Tacos Los Volcanes	1.666667	No Alcohol Served
El Rincon De San Francisco	1.666667	No Alcohol Served
Mariscos Tia Licha	1.6	No Alcohol Served
Kiku Cuernavaca	1.6	No Alcohol Served
Preambulo Wifi Zone Cafe	1.583333	No Alcohol Served
El Mundo De La Pasta	1.5	No Alcohol Served

Best Restaurants on Average for Wine-Beer

name	avg_rating	alcohol
Restaurant Las Mananitas	2	Wine-Beer
Emilianos	2	Wine-Beer
Restaurante La Parroquia Potosina	1.75	Wine-Beer
Restaurant Bar Hacienda Los Martinez	1.666667	Wine-Beer
Sanborns Casa Piedra	1.555556	Wine-Beer
La Virreina	1.533333	Wine-Beer
Restaurante Bar El Gallinero	1.5	Wine-Beer
El Cotorreo	1.5	Wine-Beer
Restaurante Guerra	1.5	Wine-Beer

Best Restaurants on Average with a Full Bar

name	avg_rating	alcohol
Rincon Del Bife	1.666667	Full Bar
El Oceano Dorado	1.5	Full Bar
Rockabilly	1.5	Full Bar

(Table 3. List of best restaurants with no alcohol, full bar, and wine/beer.)

- Based on ambiance with average rating over 1.5

Best Restaurants for Family

name	avg_rating	Rambience
Restaurant Las Mananitas	2	familiar
Michiko Restaurant Japones	2	familiar
Emilianos	2	familiar
Cafe Punta Del Cielo	1.833333	familiar
La Estrella De Dimas	1.8	familiar
Giovannis	1.75	familiar
Restaurante La Parroquia Potosina	1.75	familiar
Log Yin	1.75	familiar
La Cochinita Pibil Restaurante Yucateco	1.714286	familiar
Mariscos El Pescador	1.692308	familiar
El Rincon De San Francisco	1.666667	familiar
Restaurant Bar Hacienda Los Martinez	1.666667	familiar
Rincon Del Bife	1.666667	familiar
Mariscos Tia Licha	1.6	familiar
Kiku Cuernavaca	1.6	familiar
Sanborns Casa Piedra	1.555556	familiar
La Virreina	1.533333	familiar
El Oceano Dorado	1.5	familiar
El Mundo De La Pasta	1.5	familiar
Restaurante Bar El Gallinero	1.5	familiar
El Cotorreo	1.5	familiar
Restaurante Guerra	1.5	familiar

Best Restaurants for Quiet Environment

name	avg_rating	Rambience
Tacos Los Volcanes	1.666667	quiet
Preambulo Wifi Zone Cafe	1.583333	quiet
Rockabilly	1.5	quiet

(Table 4. List of best restaurants for different ambiance setting.)

- Based on dress code with average rating over 1.5

Best Restaurants for Informal Events

name	avg_rating	dress_code
Michiko Restaurant Japones	2	informal
Emilianos	2	informal
La Estrella De Dimas	1.8	informal
Giovannis	1.75	informal
Restaurante La Parroquia Potosina	1.75	informal
Log Yin	1.75	informal
La Cochinita Pibil Restaurante Yucateco	1.714286	informal
Mariscos El Pescador	1.692308	informal
Tacos Los Volcanes	1.666667	informal
El Rincon De San Francisco	1.666667	informal
Restaurant Bar Hacienda Los Martinez	1.666667	informal
Rincon Del Bife	1.666667	informal
Kiku Cuernavaca	1.6	informal
Preambulo Wifi Zone Cafe	1.583333	informal
Sanborns Casa Piedra	1.555556	informal
La Virreina	1.533333	informal
El Oceano Dorado	1.5	informal
El Mundo De La Pasta	1.5	informal
Restaurante Bar El Gallinero	1.5	informal
El Cotorreo	1.5	informal
Restaurante Guerra	1.5	informal
Rockabilly	1.5	informal

Best Restaurants for Casual

name	avg_rating	dress_code
Mariscos Tia Licha	1.6	casual

Best Restaurants for Formal Events

name	avg_rating	dress_code
Restaurant Las Mananitas	2	formal
Cafe Punta Del Cielo	1.833333	formal

(Table 5. List of best restaurants for informal, formal, and casual events.)

- Based on budget with average rating over 1.5

Best Restaurants for Low Budget Reviewers

name	avg_rating	avg_food
Dairy Queen	2	2
Gordas De Morales	2	2
Restaurante De Mariscos La Langosta	2	2
Rincon Del Bife	2	2
Sanborns Casa Piedra	2	2
Hamburguesas La Perica	2	2
La Cochinita Pibil Restaurante Yucateco	2	2
Tacos Los Volcanes	2	2
Mariscos El Pescador	2	1.75
Restaurante El Reyecito	2	1.5

Best Restaurants for High Budget Reviewers

name	avg_rating	avg_food
Mariscos Tia Licha	2	2
Carnitas Mata Calle Emilio Portes Gil	2	2
Restaurant Las Mananitas	2	2
Restaurante Versalles	2	2
La Virreina	2	2
El Rincon De San Francisco	2	2
Little Cesarz	2	2
Tortas Locas Hipocampo	2	2
Sanborns Casa Piedra	2	2
Gorditas Dona Tota	2	2

Best Restaurants for High Budget Reviewers

name	avg_rating	avg_food
Mariscos Tia Licha	2	2
Carnitas Mata Calle Emilio Portes Gil	2	2
Restaurant Las Mananitas	2	2
Restaurante Versalles	2	2
La Virreina	2	2
El Rincon De San Francisco	2	2
Little Cesarz	2	2
Tortas Locas Hipocampo	2	2
Sanborns Casa Piedra	2	2
Gorditas Dona Tota	2	2

(Table 6. List of best restaurants for low, medium, and high budget.)

There were many commonalities of the restaurants rated with a score of 2, which is the best overall score a restaurant can receive. Using inline views and subqueries to combine the average commonalities of the best overall restaurants. These communities consisted of overall informal attire, medium price, no alcohol served, and no smoking area at all in these restaurants. This helps consumers know what to expect if deciding to go to a restaurant with a high overall rating. The Tables below provide more information on the average qualities of the highest-rated restaurants.

Average Counts by Dress Code

Dress Code	Average Count Dress Code
formal	7
informal	21.47917

Average Counts by Price

Price	Average Count Price
high	19.5
low	14
medium	25.08696

Average Counts by Rambience Type

Rambience	Average Count Rambience
familiar	21.5098
quiet	7.111111

(Table 7. Good Restaurant Commonalities in Dress Code, Price, and Ambience)

Average Counts by Alcohol Type

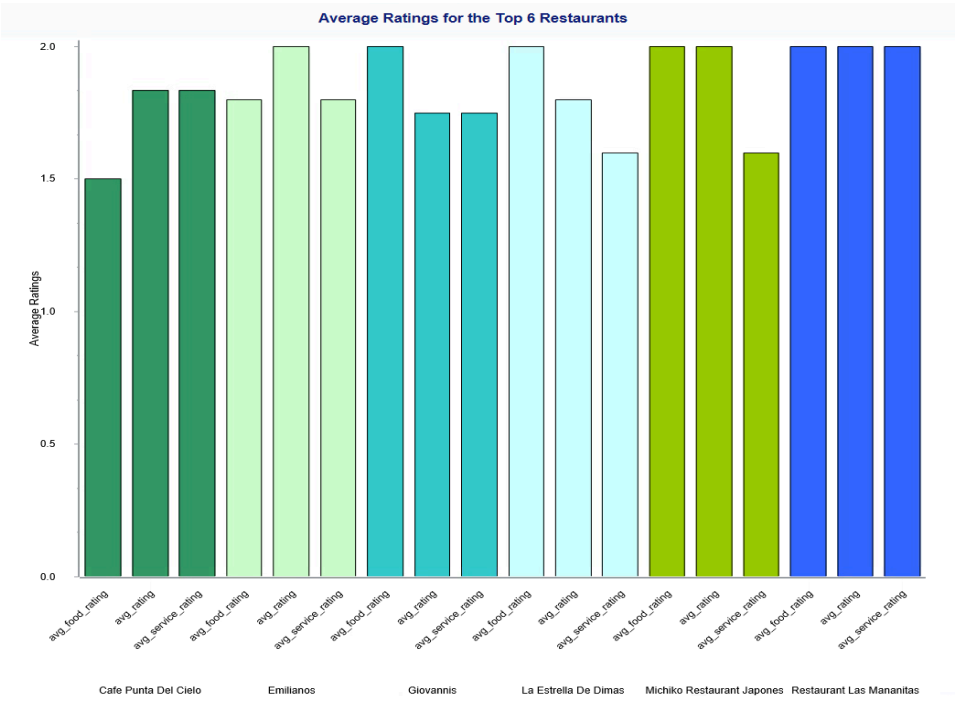
Alcohol	Average Count Alcohol
Full Bar	10
No Alcohol Served	23.09091
Wine-Beer	16.78947

Average Counts by Smoking Area

Smoking Area	Average Count Smoking Area
none	30.28571
not permitted	15
only at bar	9.5
section	17.84615

(Table 8. Good Restaurant Commonalities in Alcohol Type and Smoking Area)

The Top 6 restaurants that had an overall rating that was greater or equal to 1.75 and a food rating that was greater or equal to 1.8 is shown below in the bar graph:



(Chart 5. Average Ratings for the Top 6 Restaurants)

The bar graph above lists the six highest-rated restaurants in Mexico based on the overall rating and food rating. This helps give a visual representation of how the highest restaurants compare to one another in terms of food, service, and overall ratings. The best restaurant in terms of ratings is Restaurant Las Mananitas (ratings of 2 in each category); the second best is Machiko Restaurant Japones; however, the service rating is approx. 1.5 compared to the first-place restaurant with a service rating of 2. Restaurant Las Mananitas is the restaurant that has the best average food and service, while Machiko Restaurant Japones is the best for average food but not average service ratings.

Suggestions for Future Studies:
Since our project used a 2012 dataset, we believe if we collected data from more recent years then the findings could be different, such as different preferred payment methods. Moreover, considering the increased use of technology in restaurants nowadays, we can assume that people rate restaurants with different approaches compared to a decade ago. Therefore, for future restaurant analysis, we believe including technology such as digital wallet payment (Apple Pay/Venmo), wifi availability, and QR code menu would be more useful for both the restaurants and customers. Furthermore, we can also incorporate data collection on wait times (or other relevant numeric metrics) that could be plotted against rating to investigate any possible relationship, and take online review into consideration such as Yelp and Google.

Conclusion:
To conclude our project, we were able to give restaurants recommendations using SAS EG and cater to different customers preferences. We found that in 2012, Restaurant Las Mananitas is the highest rated restaurant in all categories(overall, food, and service), most users have medium budget for restaurants and prefer cash as the main form of payment, and lastly, the highest rated restaurant characteristics are formal setting and higher price.