In this project we wanted to create a data analytics tool that would show the popularity and emerging trends among specific specified food categories and regions. Visualizing the data allows for clear insights that are easy to understand. We used historical information to show changes over time in popularity.

In our group, more than one of us work in the food industry and a tool such as this would give us a better understanding of the industry that we are in. We all have a misunderstanding to some level of what is popular and what is not based on our own surroundings and preferences. If you and your friends all enjoy the same type of pizza, that may lead you to believe that that is the mainstream consensus when in fact your preferences are niche. We thought it would be interesting to solve the question: in a macro sense, what types of foods are preferred.

The main source of data we used for this project came from Google Trends. Google Trends gives us an unfiltered sample of search requests that were made to Google. We decided to use non-real time data to get a better idea of historical changes. The data does not take in all of the searches for these specific categories but a representative sample. Processing all of the data would be to large to process quickly. The search data is normalized to time and location. The data points are divided by the total searches in an area and time range compared to relative popularity. The data is scaled on a range from 0 to 100. The data can have errors caused by spam or automated searches/query, making Google Trends although a great metric for search activity, not a perfect mirror for popularity. Google has built in filters to minimize data errors. Duplicate searches or searches with special characters are eliminated from the data. If a search has very few people searching for it, it will appear as a zero in the data. With Google being the dominant search engine of the world, controlling 90% off the global search engine market with 5.4 billion searches a day, it is sufficient enough to represent a population and its preferences.

This information can have many use cases. Take for example a restauranteur who is looking to open a new restaurant but is not sure as to what type of food to sell or what area to open up the restaurant. Let’s say if you are in food marketing, you can have a better idea of what audiences to target and what foods are preferred in what areas.

Between the selected food categories, pizza makes up more than 50% of the data values. In America, pizza and hamburgers are in a constant power struggle for first and second place in popularity. According to Statista.com, pizza restaurant sales in the United States reached 45.59B in 2021. These figures are likely to change as major public companies are still publishing there full year reports for 2021. Keep this in mind as you look at the comparisons of food categories.

* Code review segment

I separated the data into two categories for each region: Breakdown by city and interest over time. I changed the data values from percentage to a numeric type and changed the name of the columns. Any rows that were empty were removed. All null values were dropped either manually or with a dropna() command. Extraneous symbols were removed, and the data type became consistent between the csv files. The csv files were read into jupyter lab and each of the csv files were group by the columns or specific food categories with the as index set to false. Each of the group by were followed up with the mean command. For the visualized data, it was easier to process if it only presented the sum of all the values in the sets of data for each food category. These were all saved under different variables that would be used later. The data was then printed to make sure it worked properly. To visualize the breakdown by each city, I visualized three pie charts and to see the change in interest over time, I visualized three-line plots. Here are the visualizations:

You can see that the demand for pizza has been flat showing consistent interest in the food category. Only in Tampa however do you see a decrease in interest for pizza. Sushi and vegetarian over time both grew almost parallel to each other with Tampa once again being the exception with interest in sushi staying flat overlapping with the seafood category. A piece of data that we found interesting was that demand for seafood was flat in all three regions.

To understand each category better, we investigated popular dishes in each of the food categories. We compared two dishes from each category to each other and showed the difference between the regions for the two dishes. For the sake of presentation, we are displaying one time period, but in the code, you can adjust the time period to suit your needs. Here are the visualizations.

--Explain the insights—

As you can see, the use cases for as tool like this is vast and can be very helpful depending on your needs. From a financial perspective, this data can help inform business decisions whether it is for new businesses or established businesses to know which markets to focus on. You can have a better idea of what ingredients to keep in inventory for menu changes to suit changing consumer demands. This project was worked on between Eduardo Quin, Anthony Segovia, and Clifford Charles. Each of us focused on different parts and brought it all together to complete the project.

Thank you for your time.