Analyzing Food Delivery Trends During the Pandemic

Problem Description :

The COVID-19 pandemic has brought about many changes in the way people live their lives, including how they obtain food. With many restaurants closing their doors or operating at limited capacity, food delivery services have seen a surge in demand. In this project, we aim to analyze food delivery trends during the COVID-19 pandemic using the Grubhub dataset available on Kaggle.

Objectives

The main objective of this project is to analyze food delivery trends during the COVID-19 pandemic and gain insights into how the pandemic has affected the food delivery industry. Specifically, we aim to:

- Identify the most popular types of cuisine ordered for delivery during the pandemic
- Analyze the change in demand for food delivery services before and during the pandemic
- Identify the most popular times of day for ordering food during the pandemic

Dataset

The dataset used in this project is the Grubhub restaurant data visualization dataset available on Kaggle (https://www.kaggle.com/millennialliving/grubhub-restaurant-data-visualization). The dataset contains information on over 115,000 restaurants in the United States, including their cuisine type, location, and menu items. The dataset also includes information on customer orders, such as the order date and time, order items, and delivery location.

Background Information

The COVID-19 pandemic has had a significant impact on the food industry, with many restaurants closing their doors or operating at limited capacity to comply with social distancing guidelines. As a result, food delivery services like Grubhub have seen a surge

in demand as more people turn to delivery as a safer alternative to eating out. Analyzing food delivery trends during the pandemic can provide valuable insights into how the pandemic has affected the food industry and how businesses can adapt to changing consumer behavior.

Deliverables

The deliverables for this project include:

- A detailed analysis of food delivery trends during the COVID-19 pandemic
- Data visualizations and graphs to illustrate the trends and insights discovered in the analysis
- A final report summarizing the findings and recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic.

Possible Framework :

1. **Data Exploration and Preparation**
- Import the Grubhub dataset from Kaggle and load it into a Jupyter Notebook or other preferred data analysis tool.
- Check the structure of the dataset and examine its variables to gain an understanding of the available data.
- Remove any unnecessary variables that are not needed for the analysis.
- Check for missing values, outliers, and errors in the dataset.
- Clean the data and prepare it for analysis.
2. **Exploratory Data Analysis**
- Conduct descriptive statistics to identify the most popular types of cuisine ordered for delivery during the pandemic.
- Analyze the change in demand for food delivery services before and during the pandemic using time series analysis.
- Identify the most popular times of day for ordering food during the pandemic.
- Create data visualizations and graphs to illustrate the trends and insights discovered in the analysis.
3. **Predictive Analysis**
- Develop predictive models to forecast future trends in food delivery demand.
- Evaluate the accuracy of the models using statistical measures and compare them to current trends.
- Provide recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic based on the predictive models.
4. **Conclusion and Recommendations**
- Summarize the findings from the analysis.
- Provide recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic based on the insights gained from the analysis and predictive models.
- Discuss the limitations of the analysis and suggest areas for future research.

Deliverables

The deliverables for this project include:

- A Jupyter Notebook or other preferred data analysis tool containing the code and analysis for the project.

- A final report summarizing the findings and recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic.
- Data visualizations and graphs to illustrate the trends and insights discovered in the analysis.
- Documentation of the predictive models developed for the project.

Code Explanation :

Here is the simple explanation for the code you can find at code.py file.

Data Exploration and Preparation

In this section, we are importing the necessary Python libraries, such as Pandas, Matplotlib, and NumPy, and then importing the Grubhub dataset from Kaggle and loading it into a Pandas dataframe. We are then checking the structure of the dataset and examining its variables to gain an understanding of the available data. Next, we are removing any unnecessary variables that are not needed for the analysis and checking for missing values, outliers, and errors in the dataset. Finally, we are cleaning the data and preparing it for analysis.

Exploratory Data Analysis

In this section, we are conducting descriptive statistics to identify the most popular types of cuisine ordered for delivery during the pandemic. We are also analyzing the change in demand for food delivery services before and during the pandemic using time series analysis. Finally, we are identifying the most popular times of day for ordering food during the pandemic. We are creating data visualizations and graphs to illustrate the trends and insights discovered in the analysis.

Predictive Analysis

In this section, we are developing predictive models to forecast future trends in food delivery demand. We are splitting the data into training and testing sets and creating a time series model to forecast future demand. Specifically, we are using the ARIMA (Autoregressive Integrated Moving Average) model to predict future trends in food delivery demand. We are then evaluating the accuracy of the models using statistical measures and comparing them to current trends. We will provide recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic based on the predictive models.

Running the Code

To run the code, you will need to have the necessary Python libraries installed, such as Pandas, Matplotlib, and NumPy. You will also need to download the Grubhub dataset from Kaggle and save it in the same directory as the Jupyter Notebook or Python script.

Once you have the necessary requirements, you can run the code in a Jupyter Notebook or other preferred data analysis tool.

Conclusion

The code provided is just a starting point for the project, and there are many ways to modify and expand upon it to meet the specific objectives of the project. Remember that the ultimate goal of this project is to gain insights into food delivery trends during the COVID-19 pandemic and provide recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic. By exploring and analyzing the Grubhub dataset, we can gain a better understanding of how the pandemic has affected the food industry and how businesses can adapt to these changes.

Future Work :

While the initial analysis of the Grubhub dataset provided insights into food delivery trends during the COVID-19 pandemic, there is still room for further exploration and analysis. Here are some suggestions for future work that can be done to expand upon this project:

1. **Geospatial Analysis:** One potential area of exploration is geospatial analysis to identify the areas with the highest demand for food delivery services during the pandemic. This could involve plotting the locations of restaurants and customer orders on a map to identify any spatial patterns or trends.
2. **Sentiment Analysis:** Another area of exploration is sentiment analysis to gain insights into how customers are feeling about food delivery services during the pandemic. This could involve analyzing customer reviews and social media posts to identify any common themes or sentiment towards food delivery services.
3. **Market Basket Analysis:** A third area of exploration is market basket analysis to identify the most common combinations of food items ordered for delivery during the pandemic. This could involve analyzing customer order data to identify any patterns or trends in the types of food items that are typically ordered together.
4. **Machine Learning:** Finally, machine learning algorithms could be used to predict future trends in food delivery demand and provide businesses in the food industry with recommendations on how to adapt to changing consumer behavior during the pandemic.

Implementation Guide

Here is a step-by-step guide on how to implement the future work outlined above:

1. **Geospatial Analysis:**

- Collect and clean geospatial data on restaurant locations and customer orders
- Use a mapping tool, such as ArcGIS or QGIS, to plot the locations of restaurants and customer orders on a map
- Analyze the map to identify any spatial patterns or trends in food delivery demand

2. **Sentiment Analysis:**

- Collect and clean customer reviews and social media posts related to food delivery services during the pandemic
- Use a natural language processing tool, such as NLTK or spaCy, to analyze the sentiment of the text
- Identify any common themes or sentiment towards food delivery services during the pandemic

3. **Market Basket Analysis:**

- Clean and prepare customer order data
- Use a market basket analysis tool, such as Apriori or FP-Growth, to identify the most common combinations of food items ordered for delivery
- Analyze the results to identify any patterns or trends in the types of food items that are typically ordered together

4. **Machine Learning:**

- Split the data into training and testing sets
- Develop machine learning models, such as linear regression or decision trees, to predict future trends in food delivery demand
- Evaluate the accuracy of the models using statistical measures and compare them to current trends
- Provide recommendations for businesses in the food industry to adapt to changing consumer behavior during the pandemic based on the predictive models

Remember that the future work outlined above is just a starting point, and there are many other areas of exploration and analysis that can be done to gain further insights into food delivery trends during the COVID-19 pandemic.

Exercise Questions :

1 - What types of cuisine were the most popular for food delivery during the COVID-19 pandemic?

Answer: To answer this question, we can use the following code snippet:

```
cuisine_counts = df[df['date'] > '2020-03-01']['cuisine'].value_counts().nlargest(10)

print(cuisine_counts)
```

This code will return the top 10 most popular cuisine types ordered for delivery during the pandemic. The interviewer can ask the candidate to explain how they arrived at this answer and what insights they can draw from the results.

2 - How did demand for food delivery services change before and during the pandemic?

To answer this question, we can use the following code snippet:

```
demand_before_pandemic = df[df['date'] < '2020-03-01']['date'].value_counts().sort_index()

demand_during_pandemic = df[df['date'] > '2020-03-01']['date'].value_counts().sort_index()

plt.plot(demand_before_pandemic.index, demand_before_pandemic.values, label='Before Pandemic')

plt.plot(demand_during_pandemic.index, demand_during_pandemic.values, label='During Pandemic')

plt.xlabel('Date')

plt.ylabel('Number of Orders')

plt.title('Change in Demand for Food Delivery Services')

plt.legend()

plt.show()
```

This code will plot the change in demand for food delivery services before and during the pandemic. The interviewer can ask the candidate to explain the trends in the graph and what insights they can draw from the results.

3 - What were the most popular times of day for ordering food during the pandemic?

To answer this question, we can use the following code snippet:

```
df['hour'] = pd.to_datetime(df['time']).dt.hour

hourly_counts = df[df['date'] > '2020-03-01']['hour'].value_counts().sort_index()

plt.bar(hourly_counts.index, hourly_counts.values)

plt.xlabel('Hour of Day')

plt.ylabel('Number of Orders')

plt.title('Most Popular Times of Day for Ordering Food During the Pandemic')

plt.show()
```

This code will plot the most popular times of day for ordering food during the pandemic. The interviewer can ask the candidate to explain the trends in the graph and what insights they can draw from the results.

4 - How accurate was the ARIMA model in predicting future trends in food delivery demand?

To answer this question, we can use the following code snippet:

```
from statsmodels.tsa.arima_model import ARIMA

model = ARIMA(train['hour'].values, order=(2, 1, 2))

model_fit = model.fit()

predictions = model_fit.forecast(steps=len(test))
```

```
plt.plot(test['hour'].values)

plt.plot(predictions[0])

plt.xlabel('Time (Days)')

plt.ylabel('Number of Orders')

plt.title('Forecast of Food Delivery Demand')

plt.show()
```

This code will plot the actual and predicted values for future trends in food delivery demand using the ARIMA model. The interviewer can ask the candidate to explain how they evaluated the accuracy of the model and what insights they can draw from the results.

Concept Explanation :

ARIMA stands for Autoregressive Integrated Moving Average, but you can call it AR-EEE-MA if you prefer. It's a fancy name for a time series model that's used to forecast future trends in data.

Basically, the ARIMA algorithm works by using past data to predict future data. It does this by looking at three main components: autoregression (AR), differencing (I), and moving average (MA).

Autoregression is just a fancy way of saying that the future values of a variable depend on its past values. For example, if you're trying to predict tomorrow's temperature, you might look at the temperatures from the past few days to see if there's a pattern.

Differencing is used to make the data stationary, which just means that the mean and variance of the data don't change over time. This is important because the algorithm works best when the data is stationary.

Moving average is used to model the random fluctuations or "noise" in the data. For example, if you're trying to predict the number of orders for food delivery, you might look at the average number of orders over the past few days to see if there's a trend.

Now, let's put it all together with an example. Let's say you're trying to predict the number of pizzas sold at a restaurant. You have data from the past few months, but you want to know how many pizzas you'll sell next month.

First, you would use autoregression to look at the past data and see if there's a pattern. For example, if you notice that the number of pizzas sold tends to increase on Fridays and Saturdays, you might predict that you'll sell more pizzas on the next Friday or Saturday.

Next, you would use differencing to make the data stationary. This might involve subtracting the mean or taking the first difference of the data.

Finally, you would use moving average to model the random fluctuations in the data. This might involve looking at the average number of pizzas sold over the past few days to see if there's a trend.

Once you've done all of this, you can use the ARIMA algorithm to predict the number of pizzas you'll sell next month. Of course, there's always some uncertainty in any prediction, so you'll need to use statistical measures to evaluate the accuracy of the model.

So there you have it, AR-EEE-MA in a nutshell. Just remember, if you want to predict the future, you need to understand the past!