I answered the writing questions here and on my Jupiter notebook. I am just including them here to make sure they are seen.

Given the data set, do a quick exploratory data analysis to get a feel for the distributions and biases of the data.  Report any visualizations and findings used and suggest any other impactful business use cases for that data.

 Looking at this dataset we have 5 different columns, Year, Major, University, Time, Order.

 It looks like places have preferences on what they get or how the food is seasoned, this is important to think about as we would want to try and pander to these preferences to achieve the highest sales for each location.

 Just looking at the data I dont see much correlation between what major the individual is pursuing and their order but a visualization of the data through a graph would be much more helpful here than just looking at it.

 Time can be a major factor as it appears that most orders are done between lunch time (11-13) and a few for dinner (14-18) and even less for breakfast (8-11), so it is important that we monitor when people are most likely to be at the

   food truck as to limit wasted employe time and our money.

 There looks like there might be a correlation between the year and the time when the customer orders their food, but it doesn't appear to be a major correlation when just viewing the data with my eyes.

Consider implications of data collection, storage, and data biases you would consider relevant here considering Data Ethics, Business Outcomes, and Technical Implications

1: Discuss Ethical implications of these factors

    With data collection it is our duty that we can collect only data that is needed for our operations. We shouldn't be collecting private intimate details of our customers unless it is needed for us to achieve our goals as a company.

        We need to be careful with our storage of client data, hiding it and securing it to prevent any eyes but our own from viewing it. A company needs the customers’ trust for continued survival and as such we must do everything in our power to protect our client’s information at every stage of collection and when we store it ourselves. We also must refrain from biases with collection of our data, we must handle all data without any preference of the collector. We cannot handpick data from the datasets, it must be looked at as it is, and our models must reflect the real data to provide accurate results. We must also balance all of this with the success of our company as we need this data to remain competitive and to keep the company alive, but we cannot hurt the customer who we collect the data from.

2: Discuss Business outcome implications of these factors

    It is important that we consider every step we take as a company to remain unbiased in our collection and use of the data that we collect. We cannot allow an individual to have an influence on the datasets that we collect or use as it will impact the service quality we can provide for our clients. If our data collection is not handled correctly, it can have serious consequences and repercussions for our company. We must do our best to handle all aspects of data collection, storage and data biases to protect the company while providing a strong service.

3: Discuss Technical Implications

    With data collection it is important we collect only the data we need. Unless it is absolutely needed for our operations, we shouldn't hold the private information of our clients. Here with this dataset, it was handled well, we only have very broad data which prevents hostile entities from using this data in malicious ways. We can enhance the quality of our storage through encryption or obfuscation, by using code word or abbreviations or by hiding everything in a private format. We can further limit the potential damage if this data were to ever leave our hands. Data biases pose a huge threat to the practical use of the data we collect. If our data is being used to predict the future outcomes of clients who use our service, it is important that we know their true actions and not one’s cherry picked by us. Failure here would result in a predictive model that would be almost completely useless when deployed in the real world.

To determine if this is a suitable course of action, I would want to make multiple more classification models to find the one with the highest accuracy. I would also like a slightly larger dataset near around 10,000 entries so that we can maximize the accuracy of our models. I would then like to do a soft deployment where we have people just test it and compare it to the real-world results before we completely started relying on our models, just so if anything goes wrong, we have time to readjust and fix the models. Overall, I think modeling is perfect for this application, it is just finding the correct one to use.