**Question 1 (10 Points)**

List as many use cases for the dataset as possible.

1. Combine customer code with order id and item code to create unique identifier and predict which combination would churn with ‘is\_churn’ as target variable
2. Just take customer code and predict which customer would likely churn with ‘is\_churn’ as target variable
3. Combine customer code with order id and item code to create unique identifier and predict the total price
4. The above three can be done for seller replacing ‘customer\_code’ with ‘seller\_code’

**Question 2 (10 Points)**

Pick one of the use cases you listed in question 1 and describe how building a statistical model based on the dataset could best be used to improve the business this data comes from.

1. I pick point 1 from above question since there were only 838 unique customers in the dataset
2. Since we can predict whether the specific combination of customer and item is likely to churn, we can avoid marketing those products to customers and market the rest of the products which are not likely to churn for customer-item combination

**Question 3 (20 Points)**

Implement the model you described in question 2, preferably in Python. The code has to retrieve the data, train and test a statistical model, and report relevant performance criteria. Ideally, we should be able to replicate your analysis from your submitted source-code, so please explicit the versions of the tools and packages you are using.

* Please check the attached code

**Question 4 (60 Points)**

A. Explain each and every of your design choices, you can use jupyter notebooks. (e.g., preprocessing, model selection, hyperparameters, evaluation criteria). Compare and contrast your choices with alternative methodologies.

* Compared logistic regression with ada-boosting

B. Describe how you would improve the model in Question 3 if you had more time.

* I would need out of time testing data to check for over fitting problem
* I would have used deep learning like CNN, LSTM to see whether its better than traditional methods
* Do more descriptive statistics analysis using graphs and charts to find interesting trend