**Project Report**

**Title: In Vehicle Coupon Recommendation System**

Submitted by:

**J v m s s n Kapardhi [Reg. No: BL.EN.U4CSE19056]**

**K Syam sai [Reg. No: BL.EN.U4CSE19070]**

**G s s Vignesh [Reg. No: BL.EN.U4CSE19044]**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

Submitted to:

**Dr. Kumaran U**

**Dr. B. Uma Maheswari**

**November 2021**

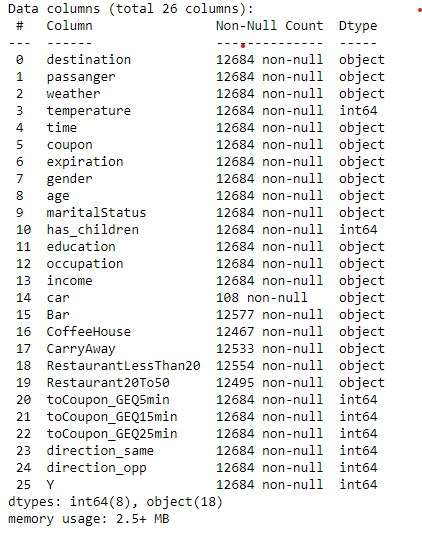
**Index:**

|  |  |
| --- | --- |
| **Page No** | **Topic** |
| 3-4 | Introduction |
| 5-17 | Implementation, Visualization, and Analysis |
| 18-20 | Machine Learning Algorithms |
| 20 | Conclusion |
| 21 | References |

**Introduction**

As technology throughout the world is evolving and lifestyles around the world are also gradually changing to adapt to different things and different sectors. In recent times, Industries are booming up in one such sector which gained high popularity is food industries (like restaurants bar, coffeehouse, etc) so to attract the customers one technique which is used by those type of industries are discount coupons offered to their customers but the main challenge arises when offering correct coupons to the customers So we've chosen a data set with features such as age and income. the frequency of the customer visiting a bar, coffee shop, or restaurant. Data also contains the gender and other personal info of customers. Data also contains coupons offered to that customer and other characteristics. As a result, it qualifies as a good data set for our issue statement, and we chose it.

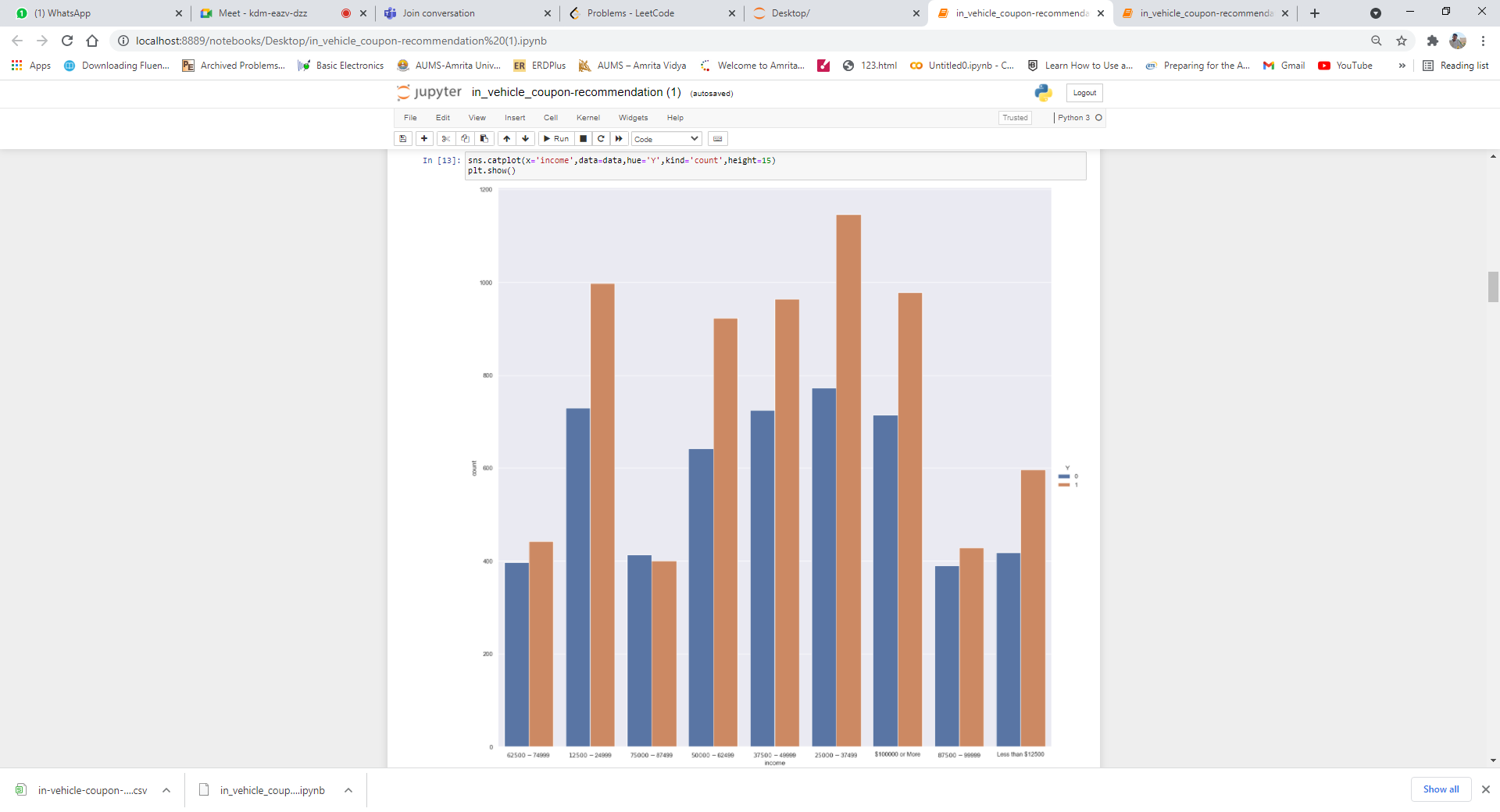
Attributes:



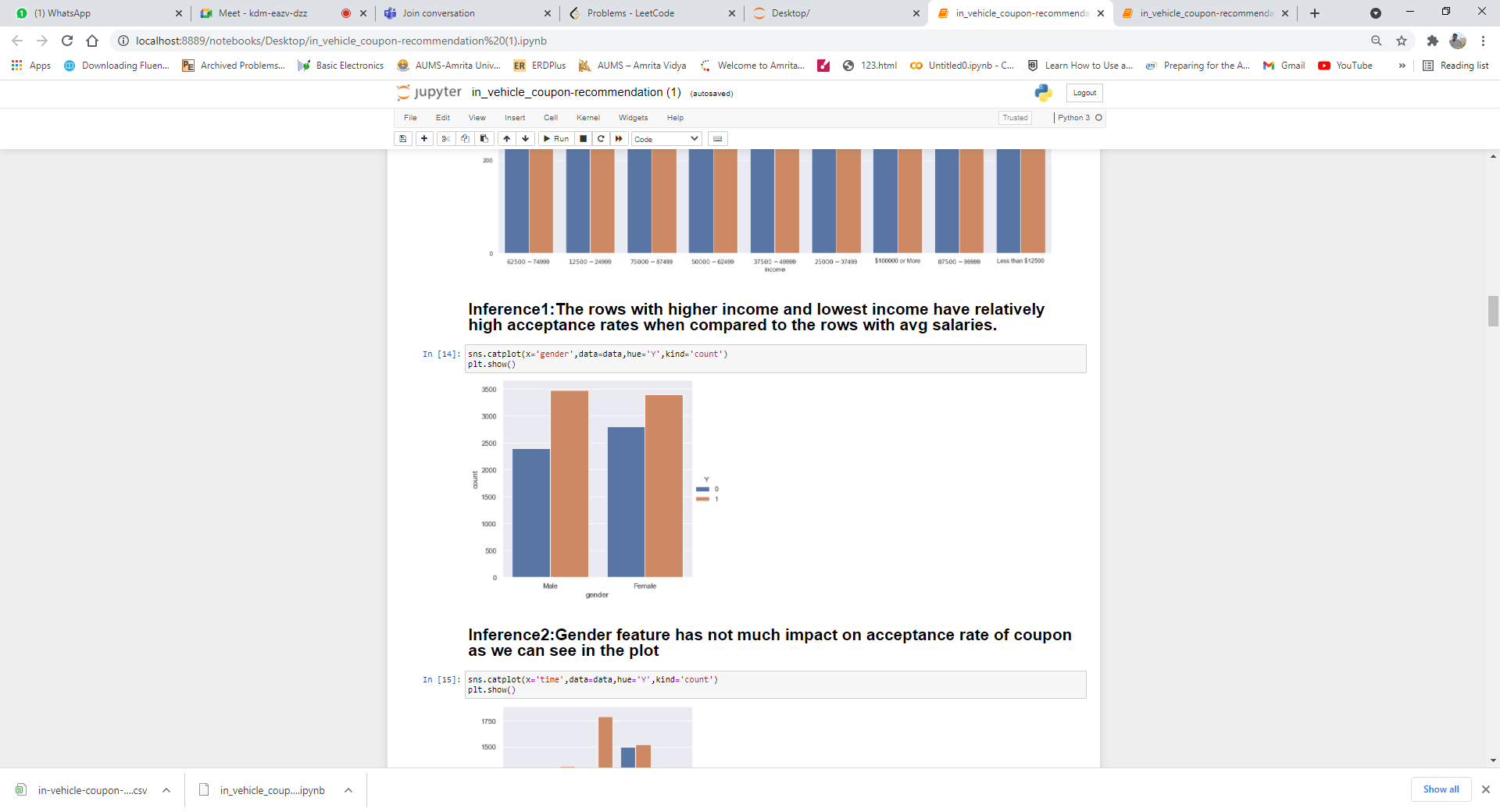
**Implementation:**

We are using python to implement our project. We are using matplotlib, seaborn inbuilt functions in python to plot various graphs and understand the relationship between the attributes. We have 26 attributes, 12685 rows in our dataset after pre-processing our initial dataset. We are applying different machine learning algorithms which are recommended for using the categorical dataset. almost all the features in the data set are categorical. so, we have encoded the features and applied the following algorithms: Naive Bayes, support vector machine, KNN, Decision Tree, Logistic regression, Tensor Flow(giving a try)

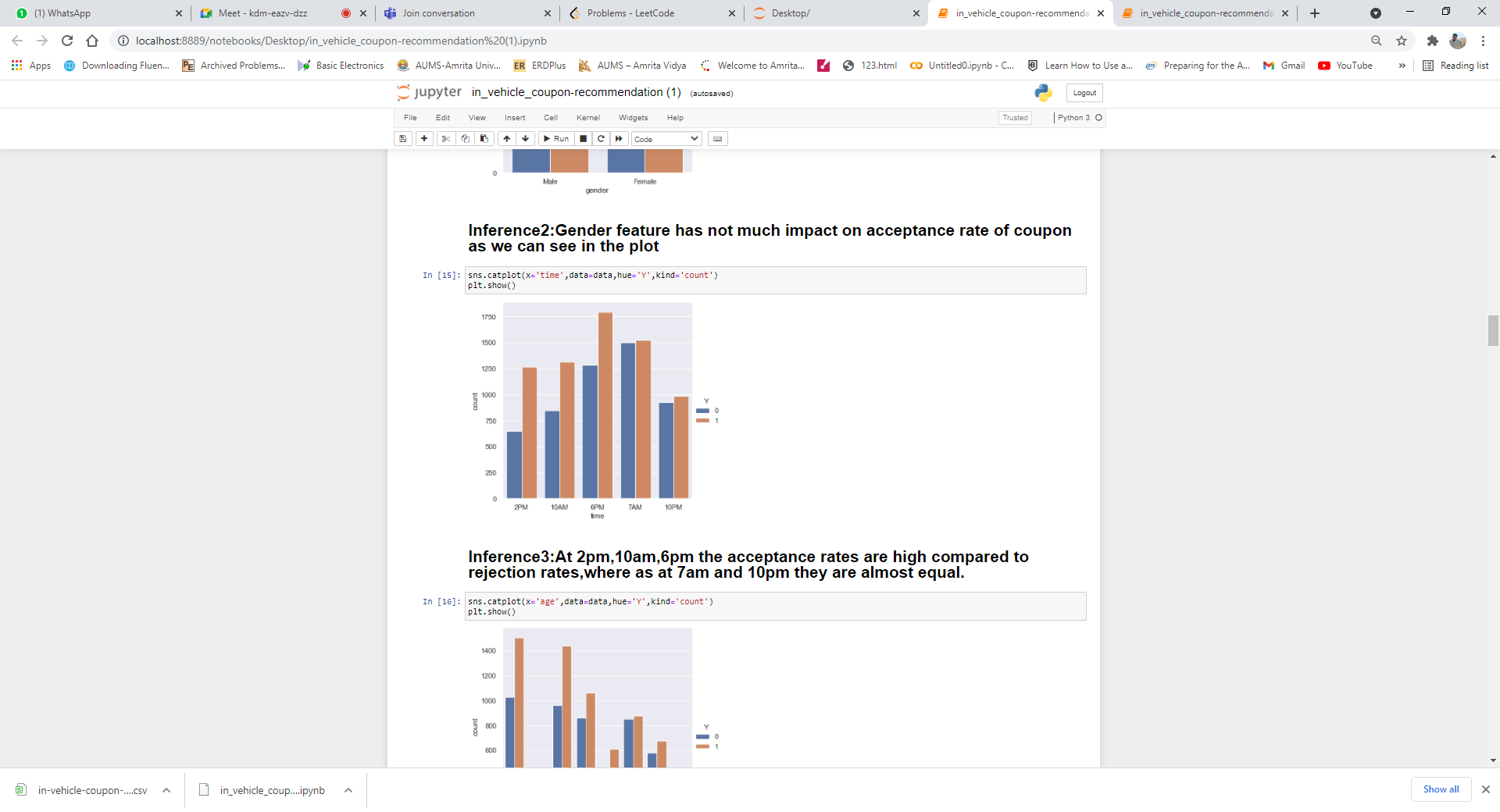
**Visualization and Analysis**



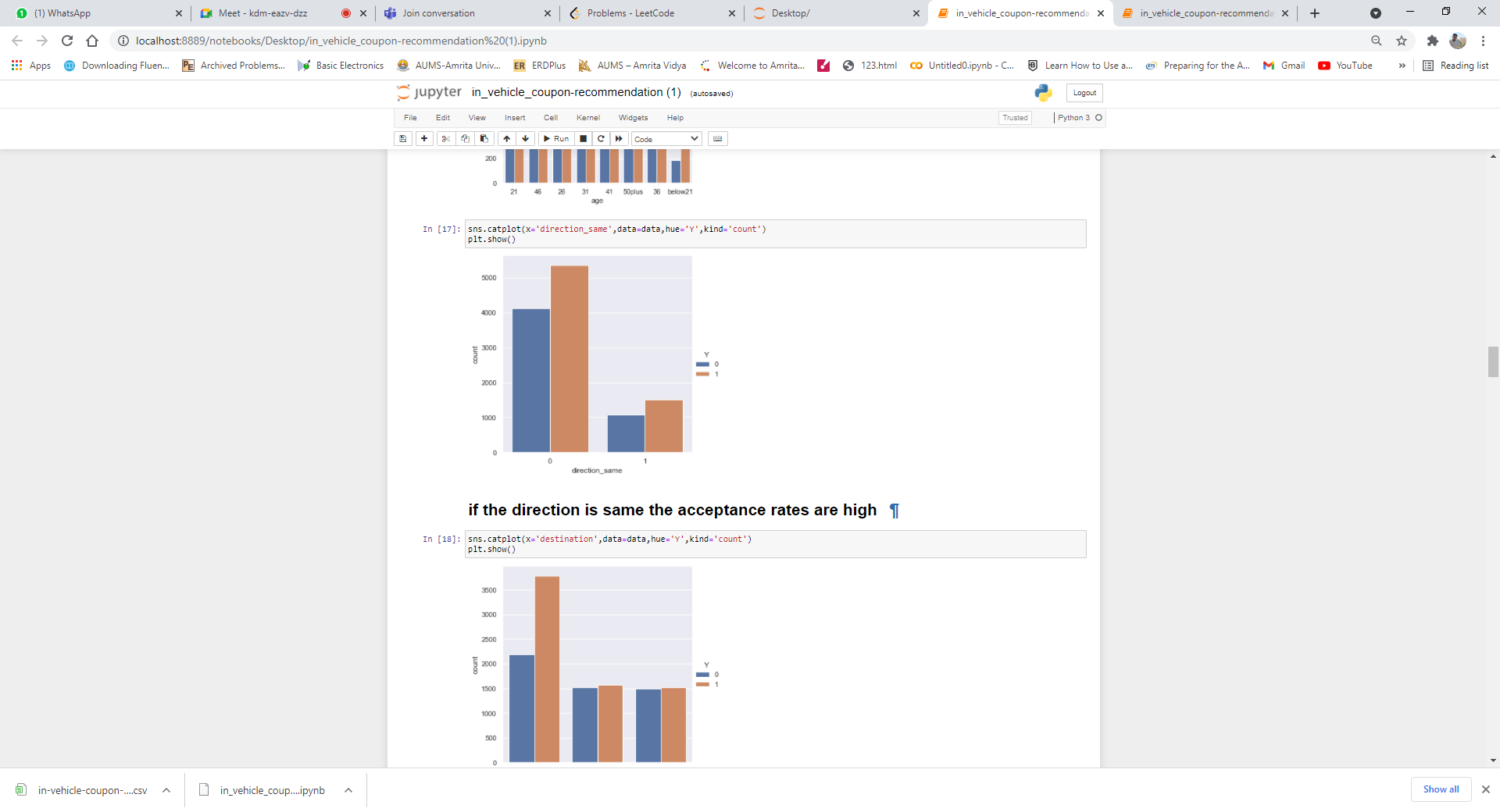
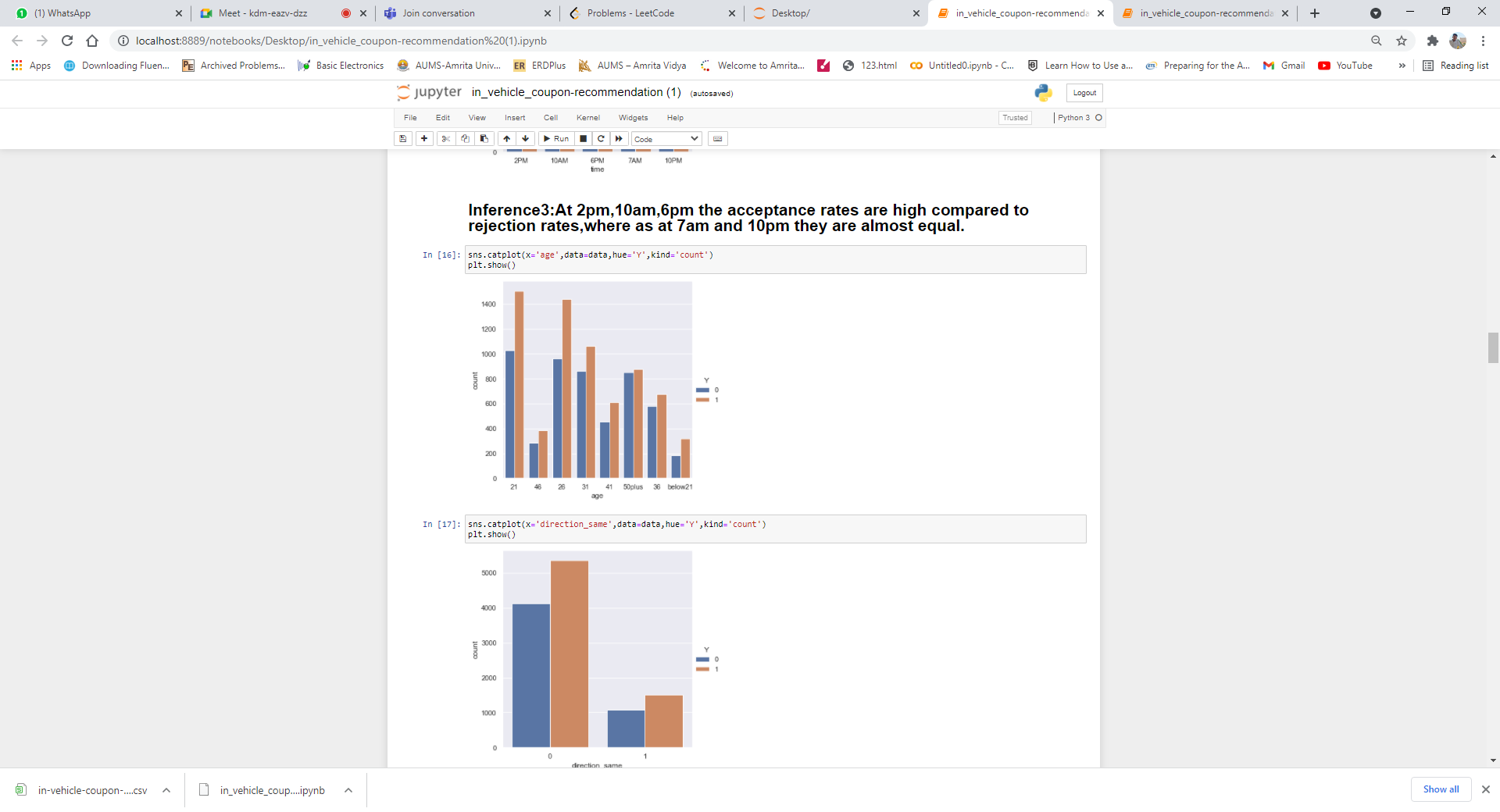
Inference: The rows with higher income and lowest income have relatively high acceptance rates when compared to the rows with average salaries.



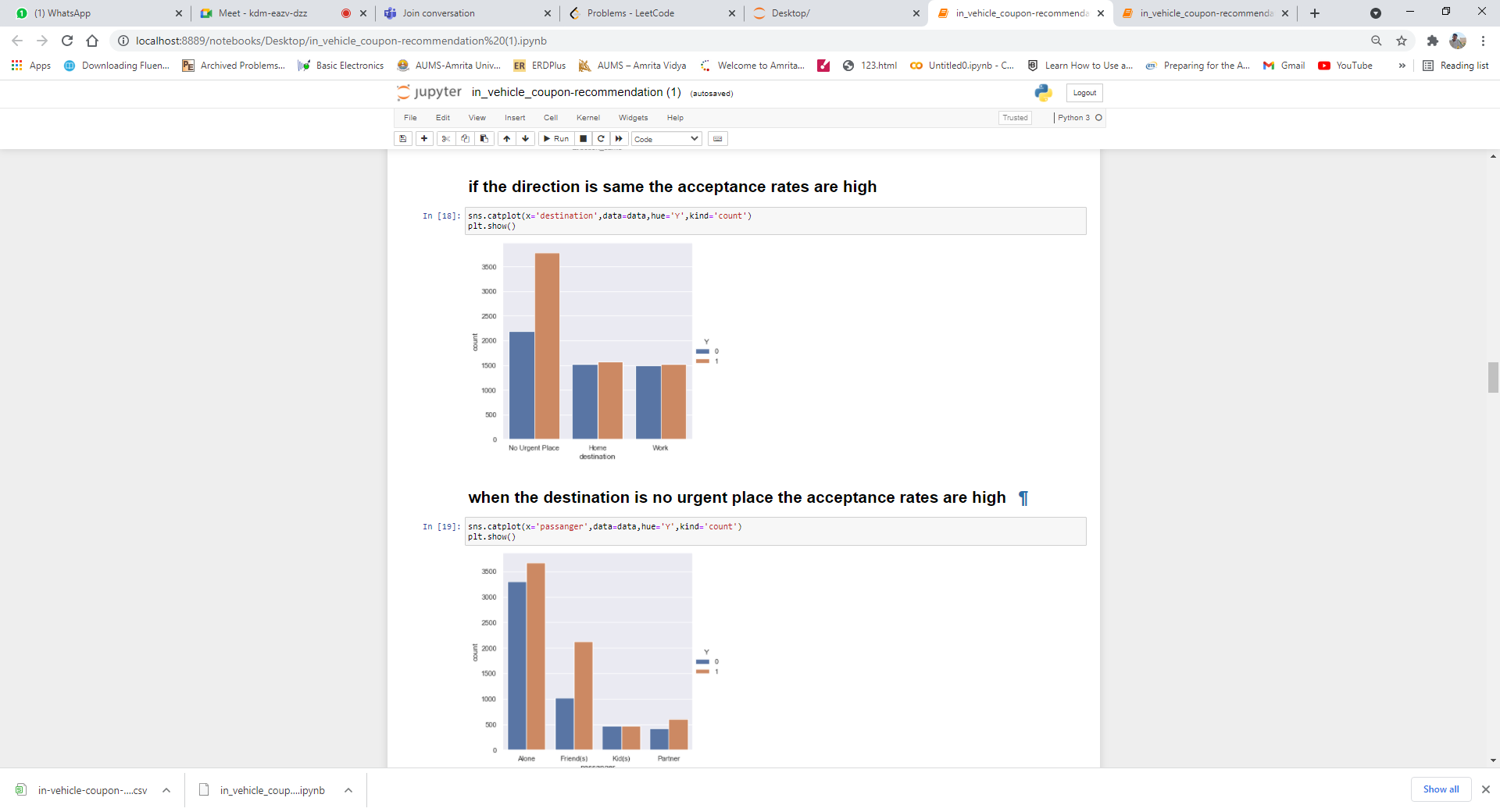
Inference: Gender feature has not much impact on the acceptance rate of the coupon as we can see in the plot



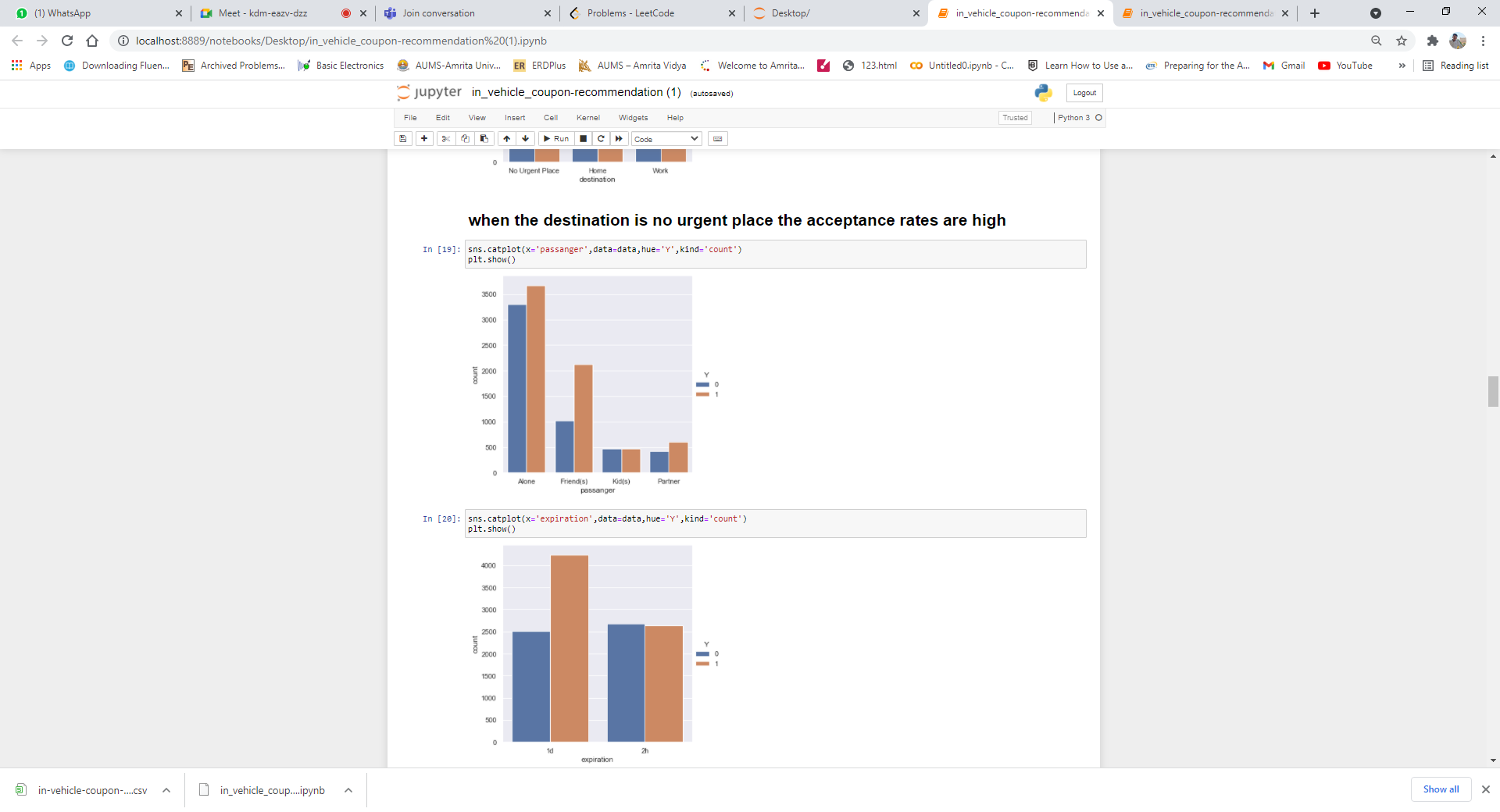
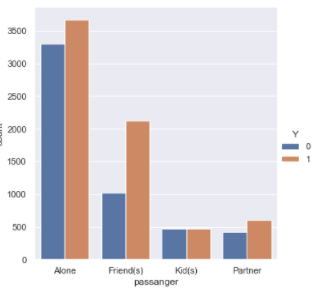
Inference: At 2 pm, 10 am, 6 pm the acceptance rates are high compared to rejection rates, whereas at 7 am and 10 pm they are almost equal.



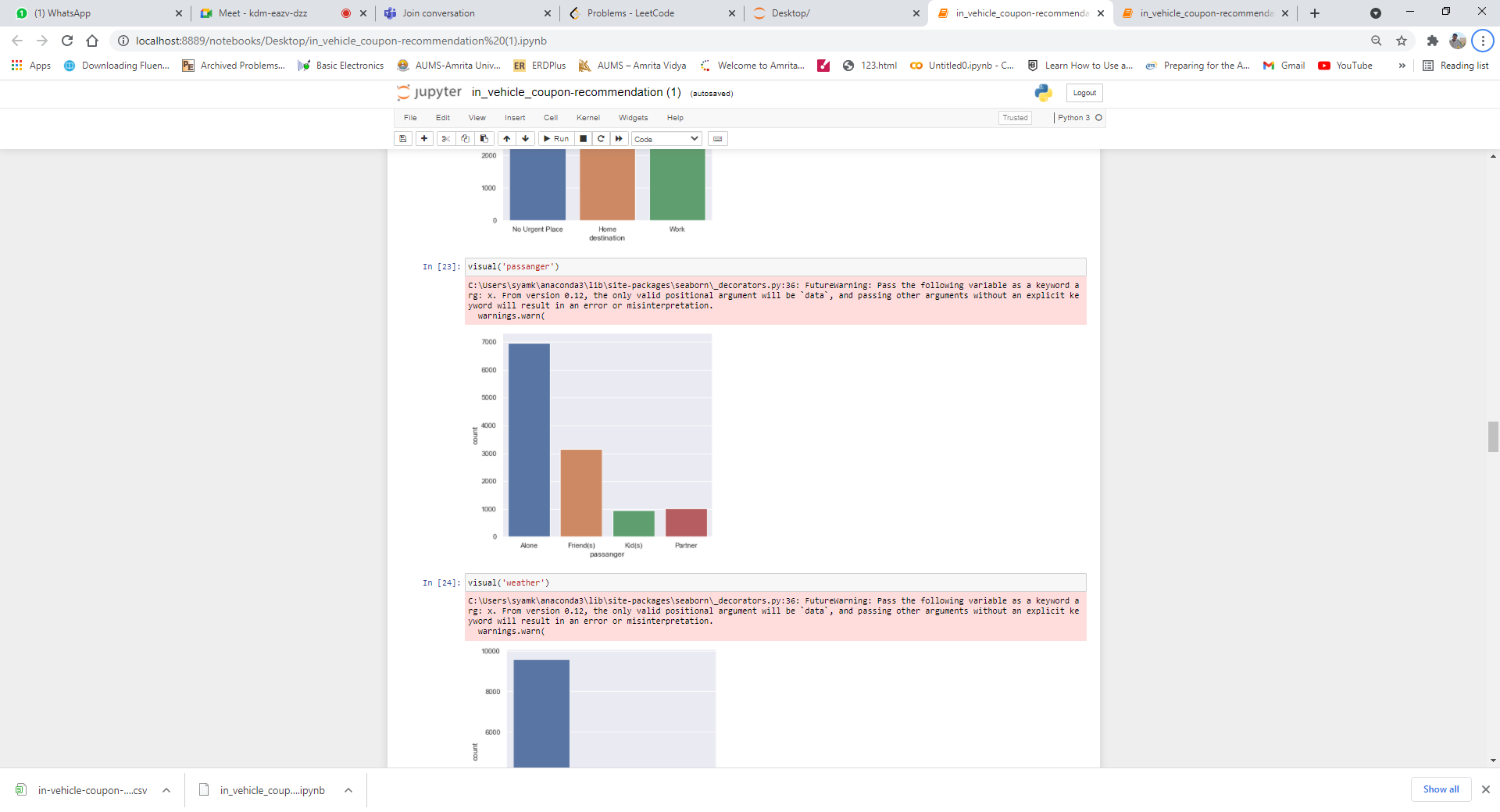
Inference: the direction is the same the acceptance rates are high

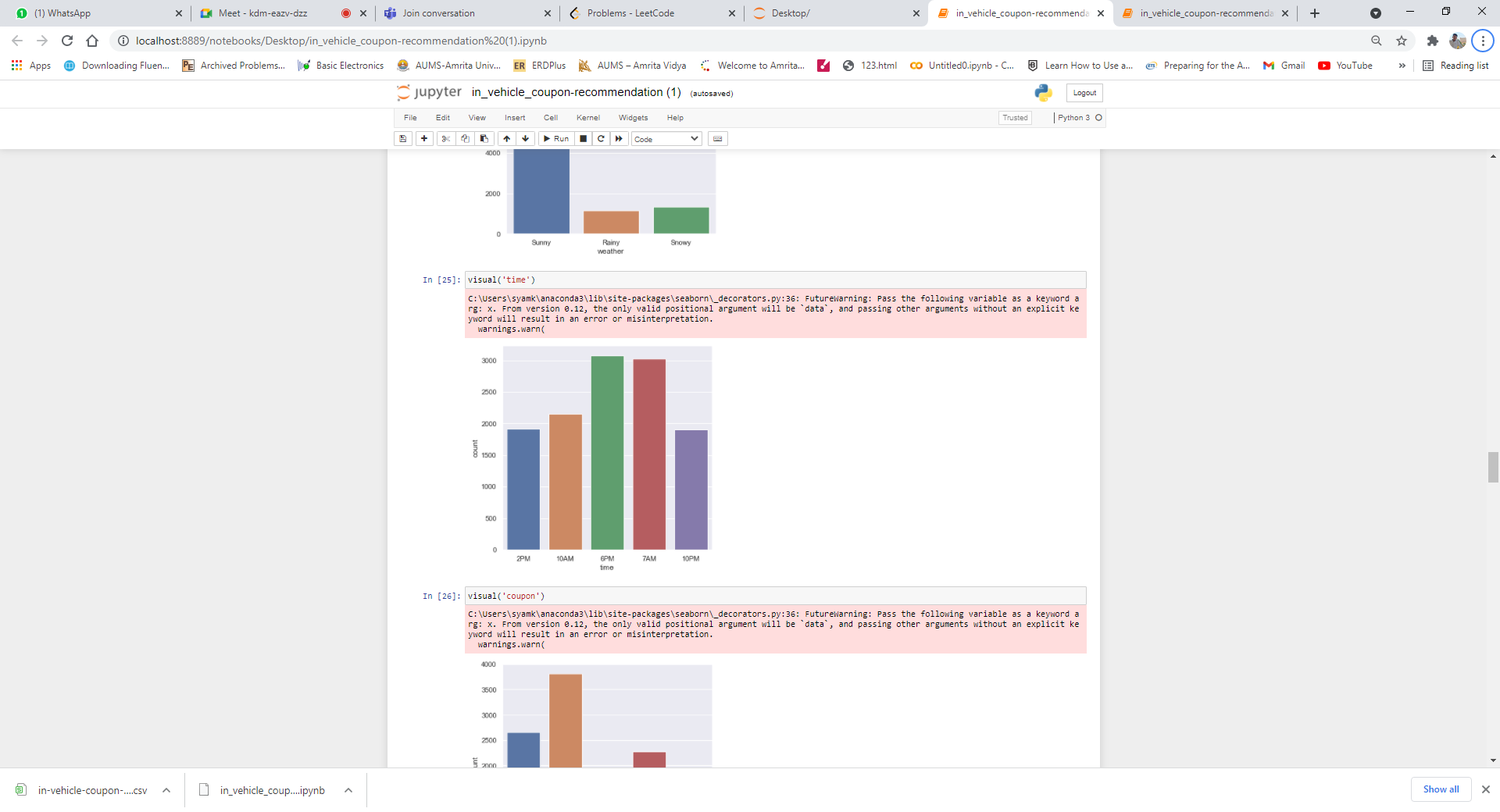
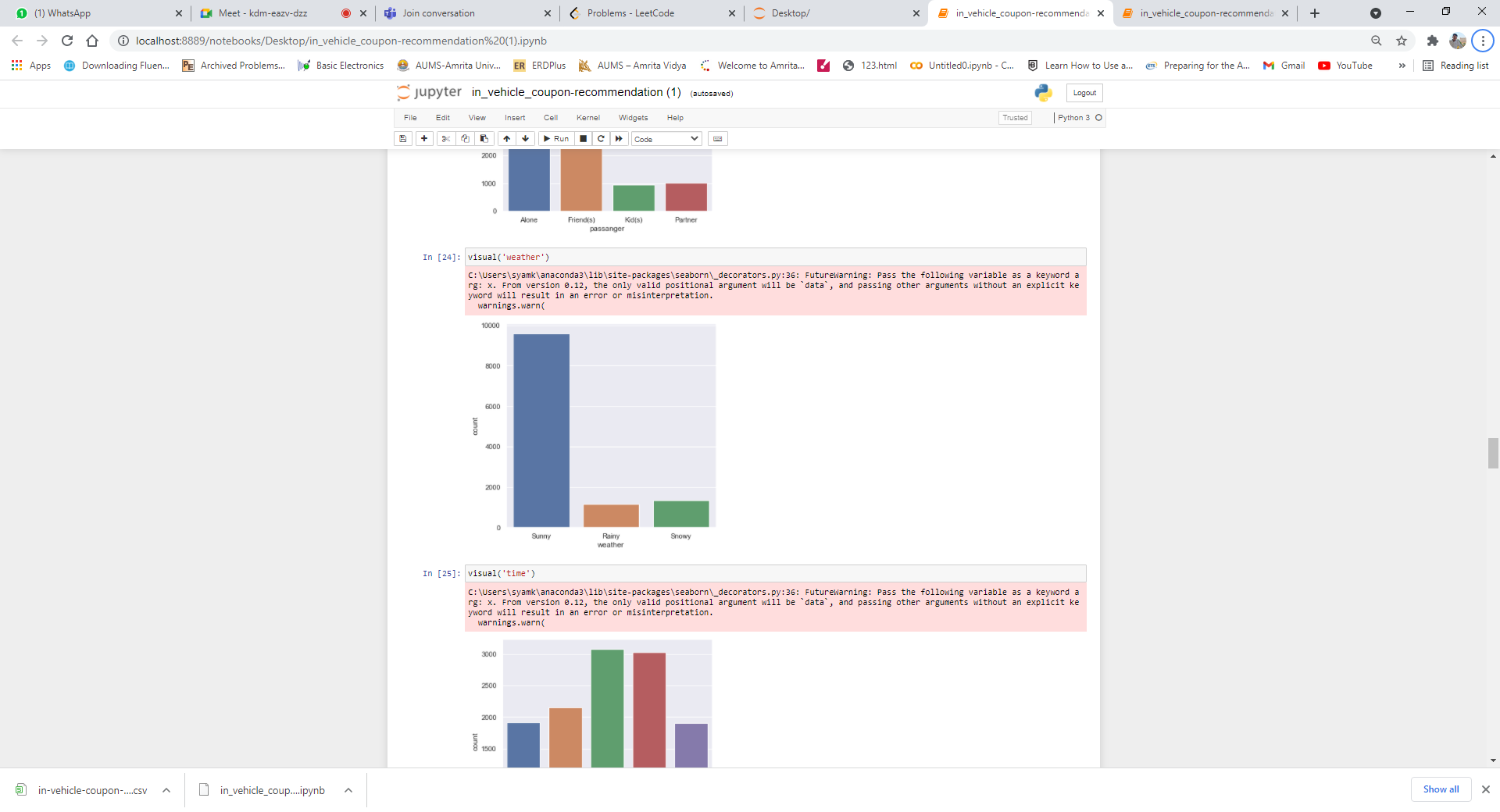


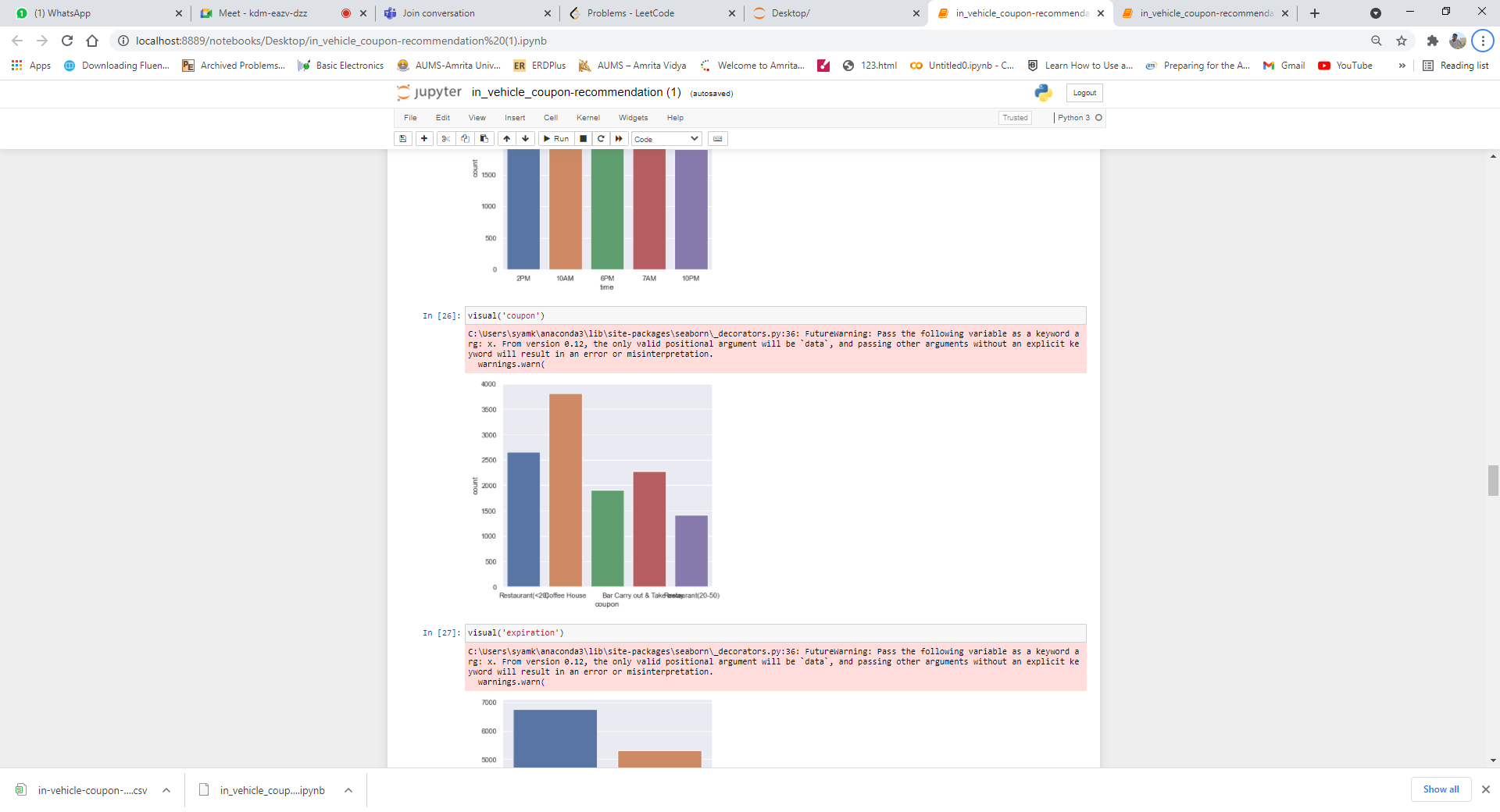
# when the destination is no urgent place the acceptance rates are high

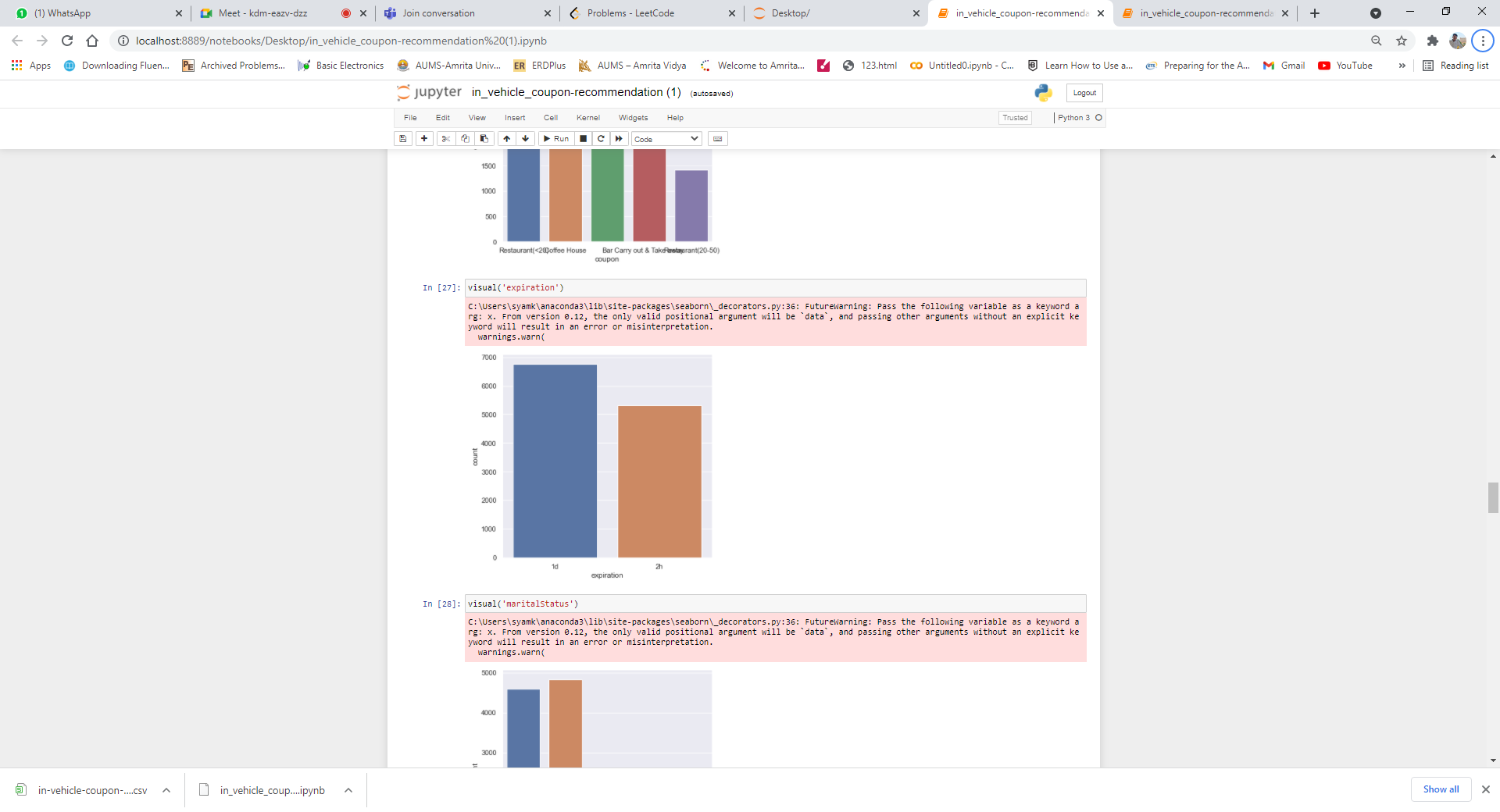


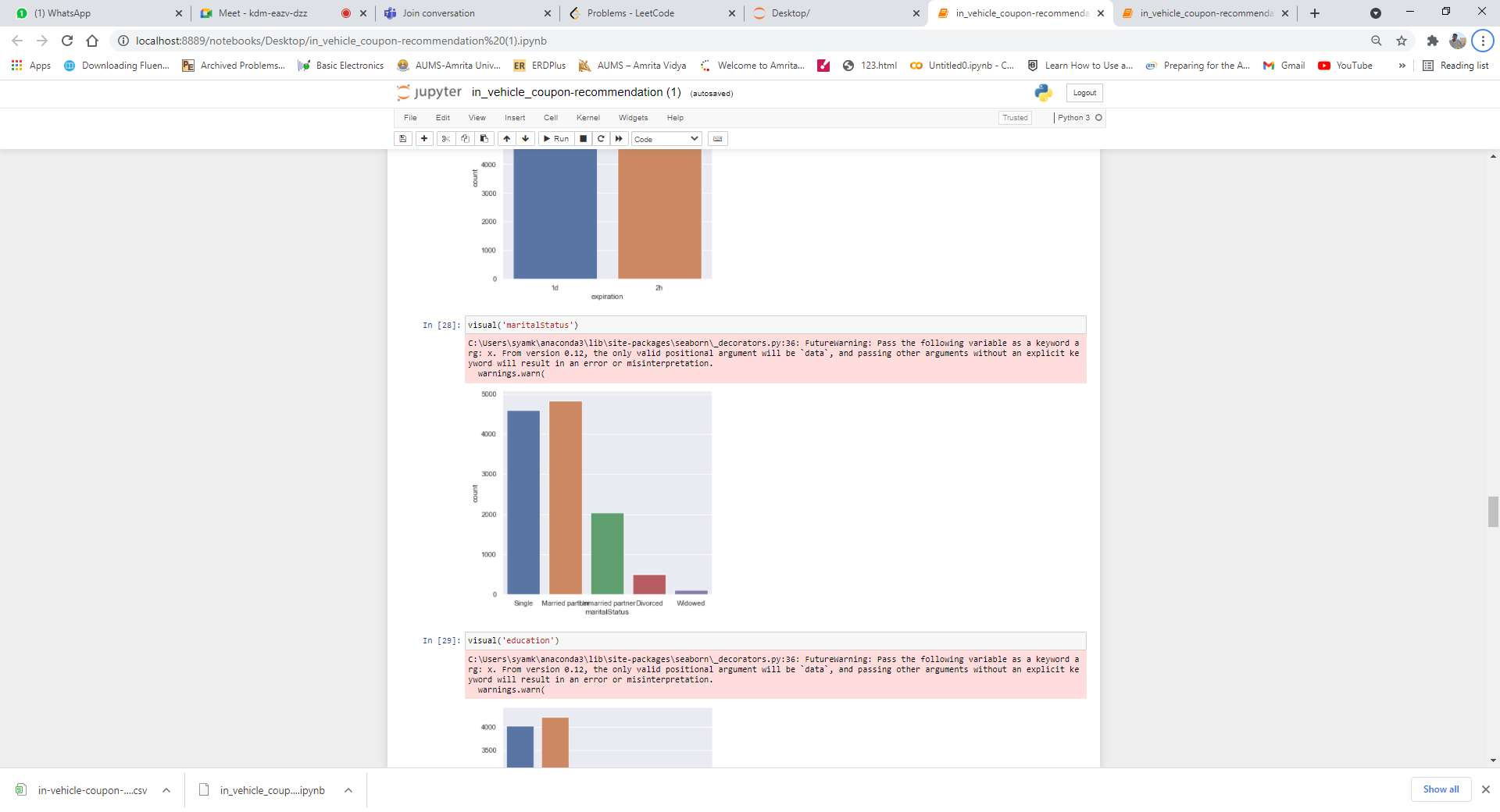
when the expiration of the coupon is offered in 1d the acceptance rates are high

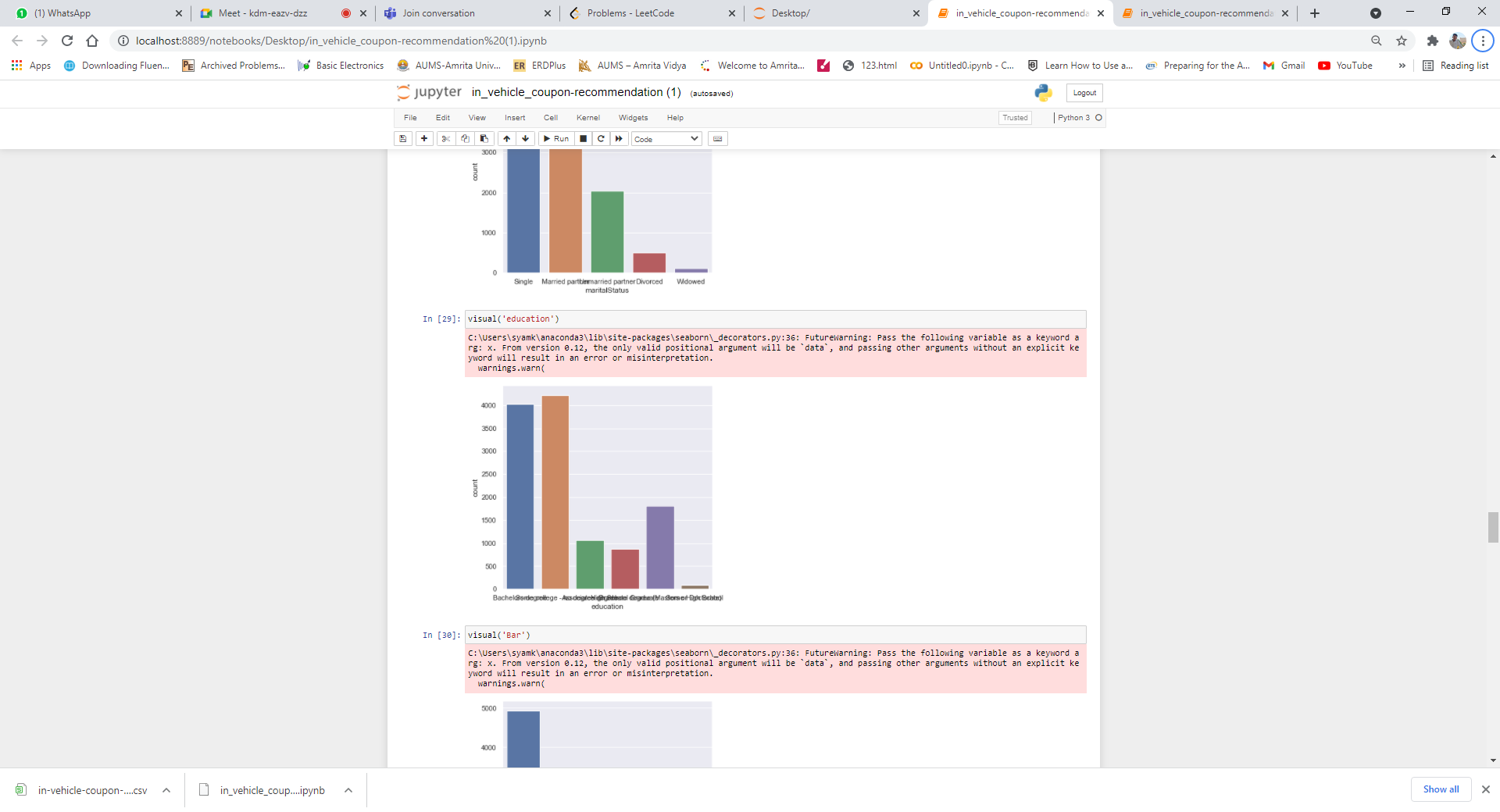
’

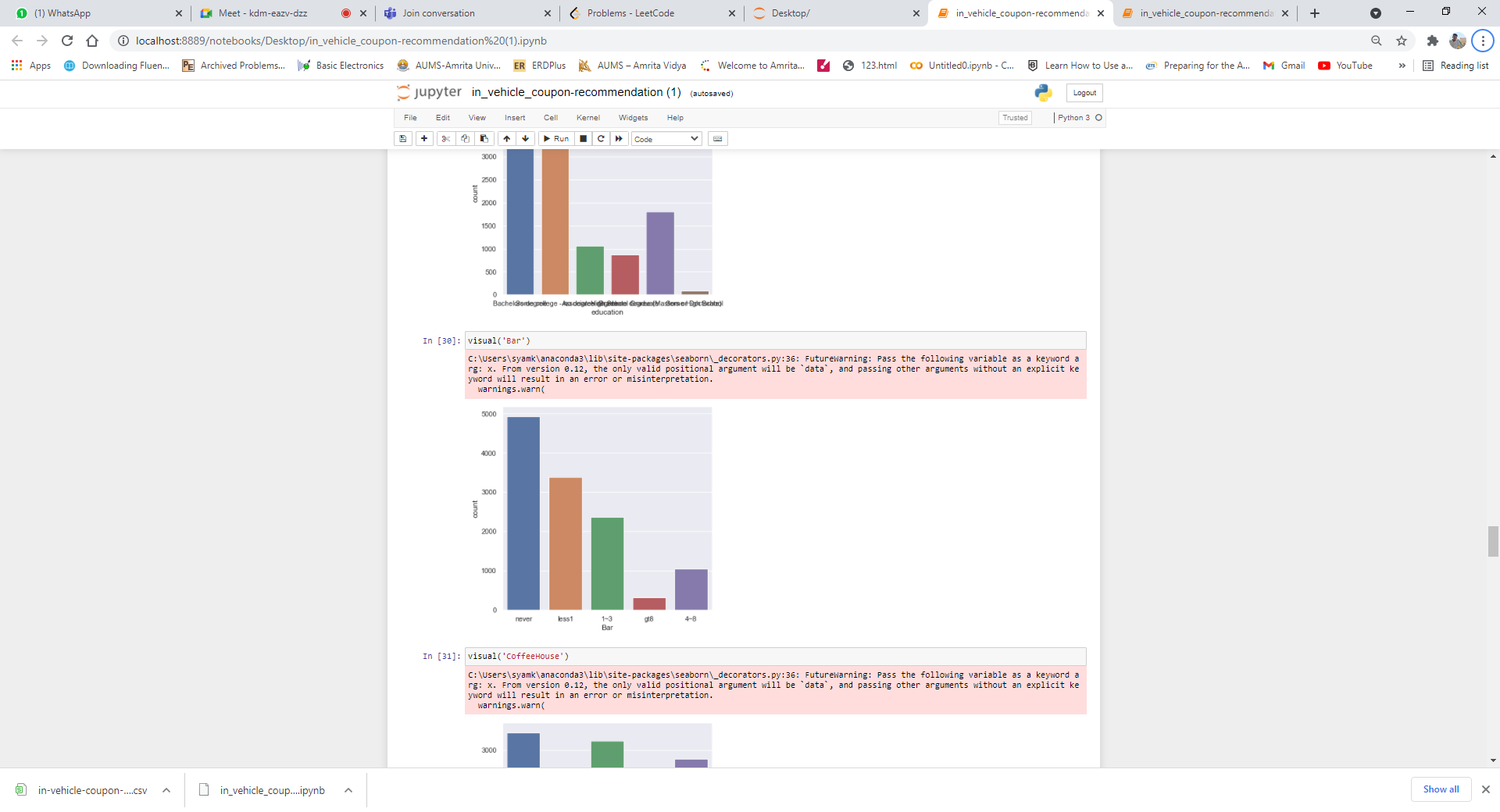


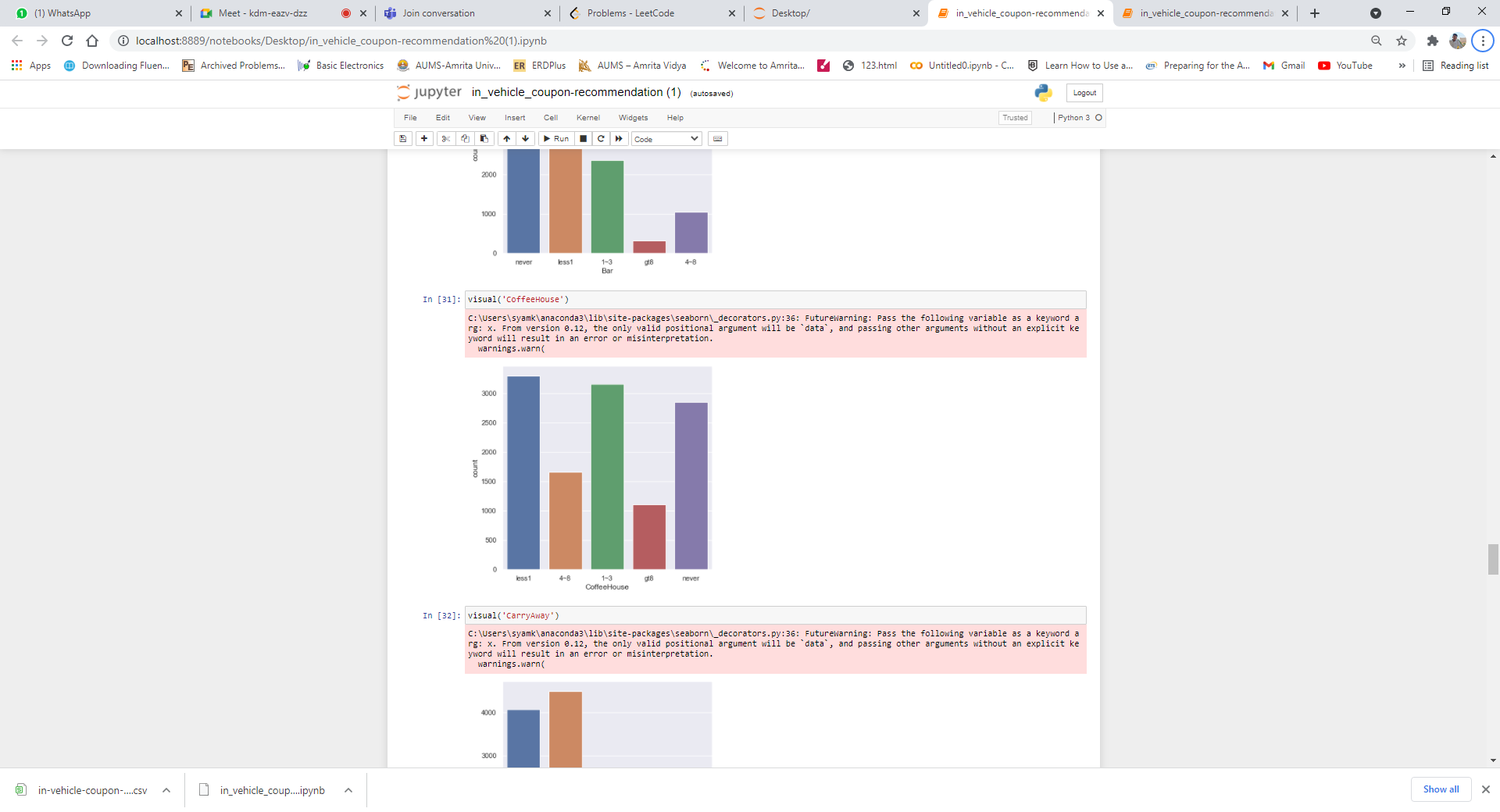


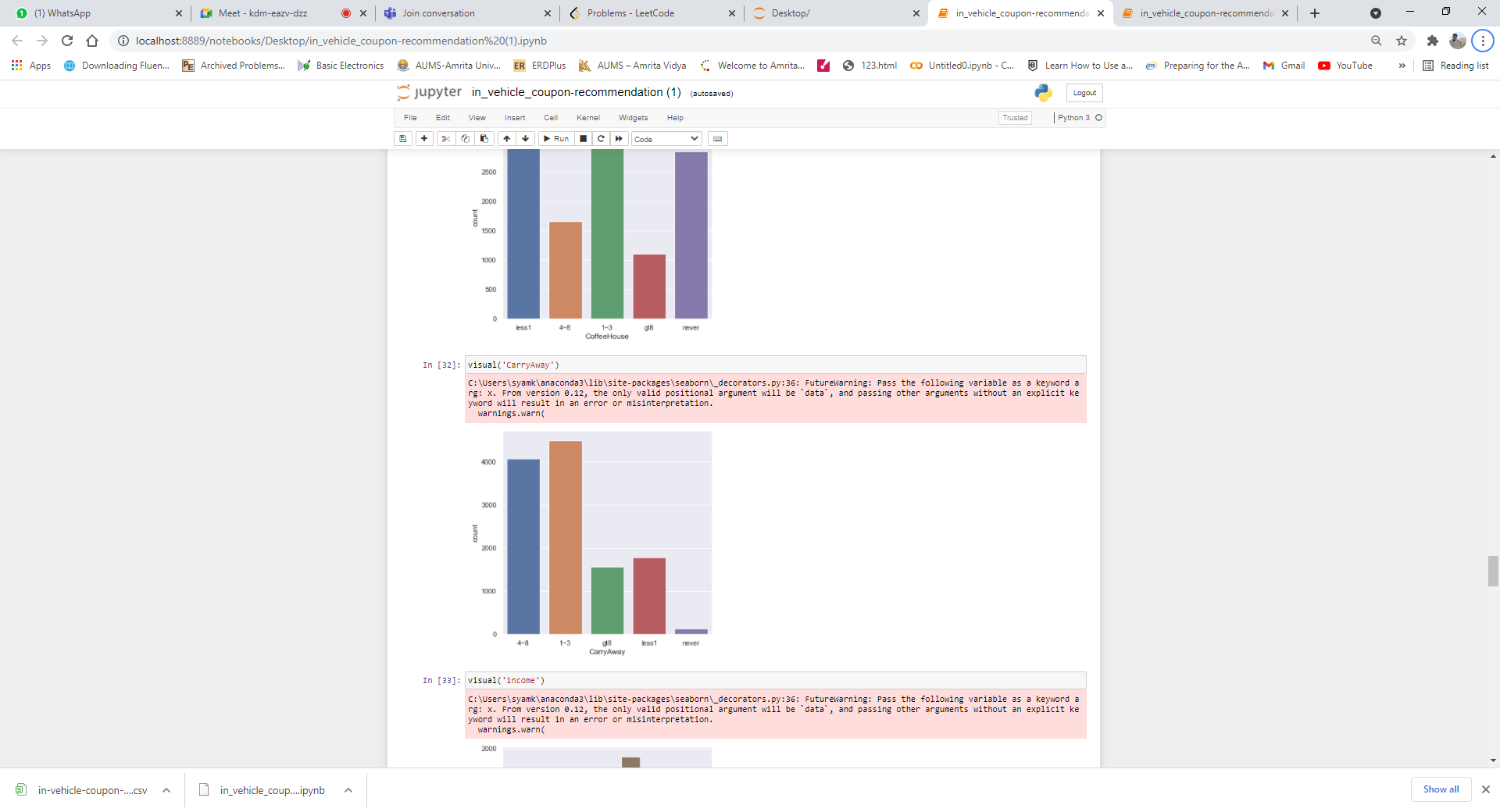


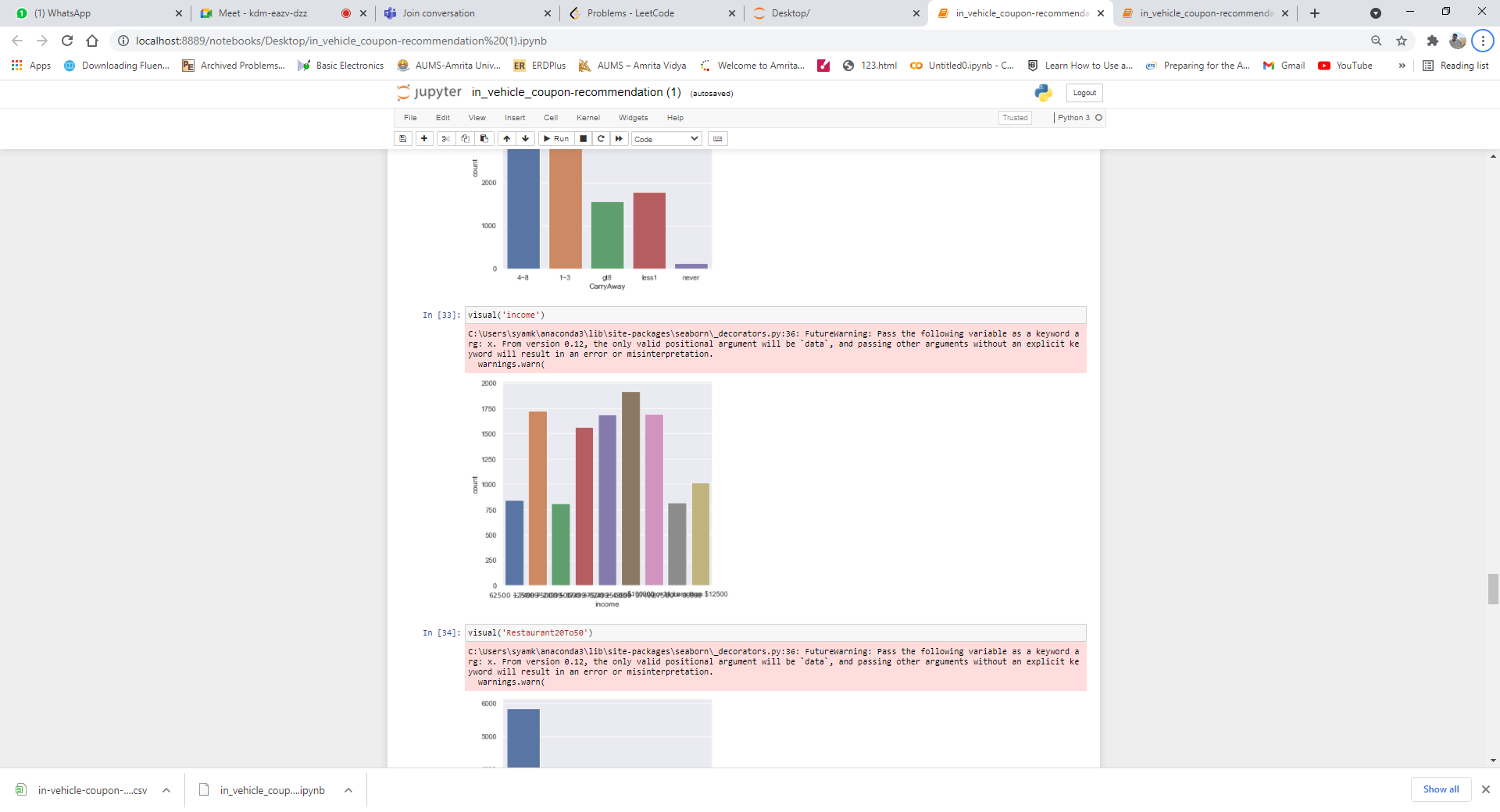


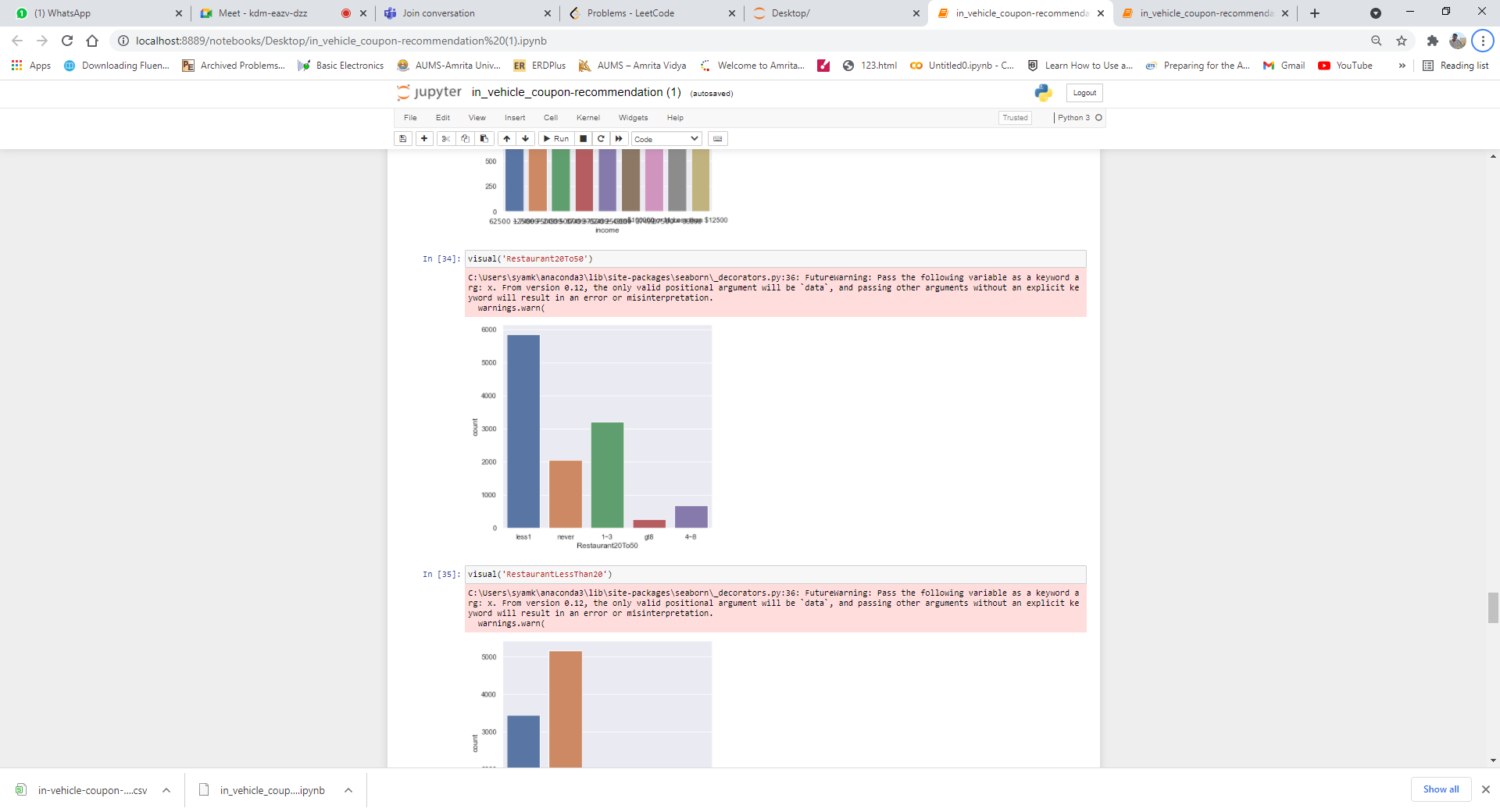


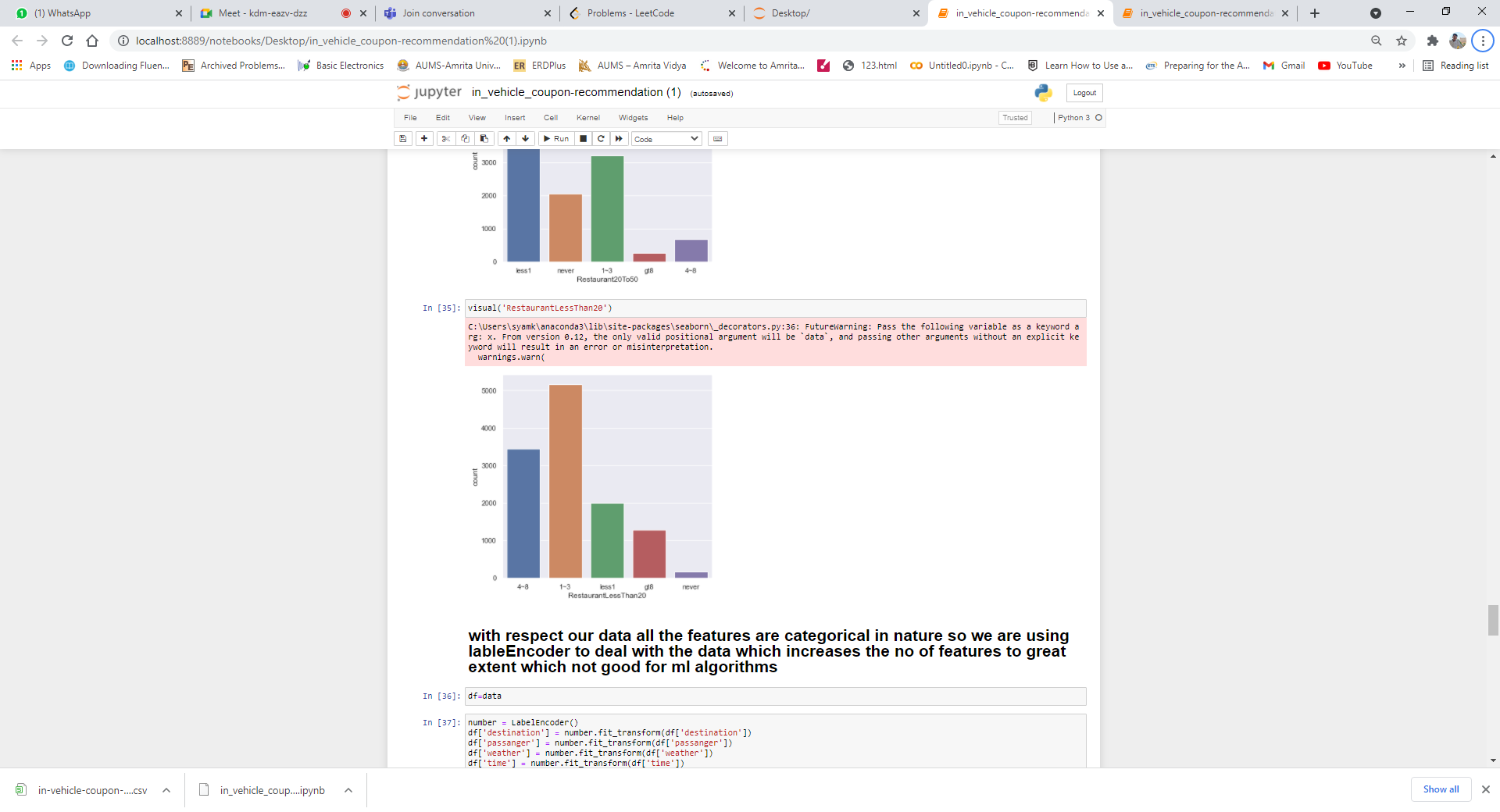










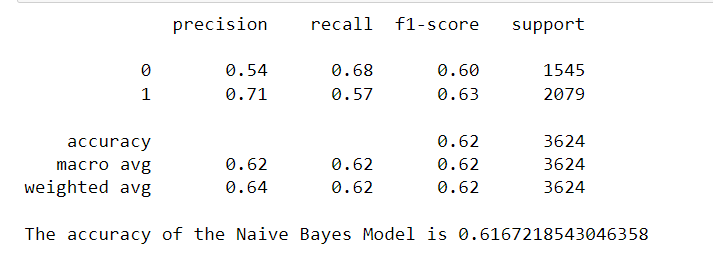


Inference: The features in our data are categorical. So, we are using the One Hot Encoding Technique.

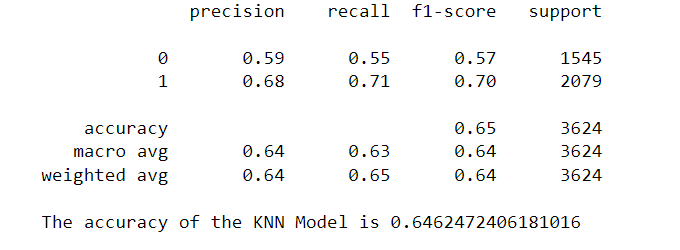
**MACHINE LEARNING ALGORITHMS AND RESULTS OBTAINED:**

As the features and targets in our dataset are categorical, we are using label encoding techniques to convert them into numerical data in pre-process phase. After that, we normalized the data to overcome problems of the dominance of one column over other because of different scales. For example, in our data set temperature feature is dominating the other columns after label encoding the categorical features. After the normalization of the data, we divided the dataset into train and test in the ratio of 80:20. We have started with linear regression. We imported the stats model to analyze whether our data is linear or not. after observing r-square, adj-r-square values, and p values. We have decided not to apply the linear regression

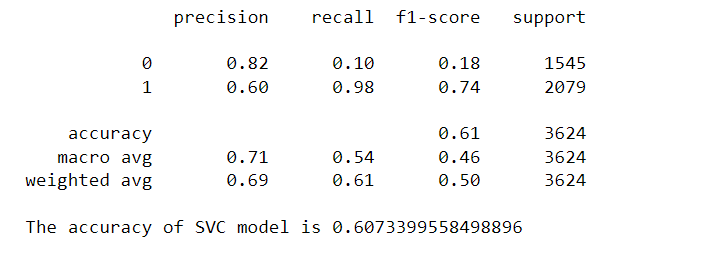
NAIVE BAYES MODEL:

****

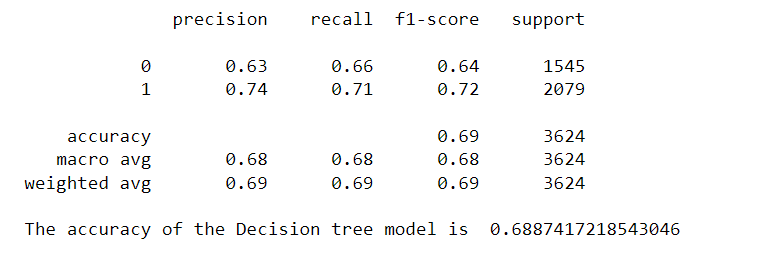
KNN MODEL:

****

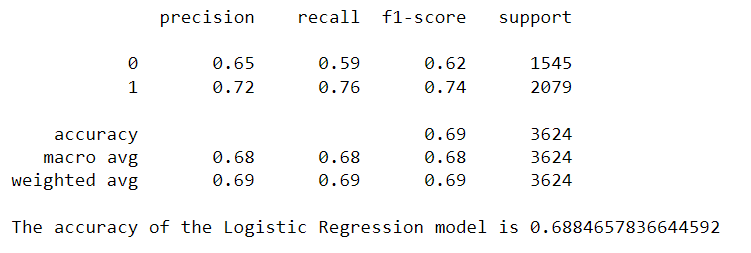
SVC MODEL:

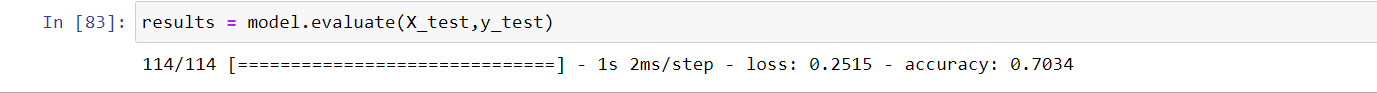
****

DECISION TREE MODEL:



Logistic Regression model:

****

TensorFlow: 

**Conclusion**

From the above classification reports, we can conclude that the logistic and decision tree models are better when compared to the other ml models for our data set with an accuracy of 68%. We have given a try with the tensor flow for which we three are currently in the learning stage, but it gave us 70% accuracy.

**References**

Visualizations:

<https://plotly.com/python/line-charts/>

<https://plotly.com/python/pie-charts/>

<https://plotly.com/python/box-plots/>

Machine Learning Algorithms:

<https://scikit-learn.org/stable/auto_examples/text/plot_document_classification_20newsgroups.html?highlight=navie%20bayes>

<https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html?highlight=svc#sklearn.svm.SVC>

<https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html>

<https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html>

[https://scikit-  
learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html](https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html)

Tensor flow:

<https://www.tensorflow.org/>