**Food & Beverages Spend Prediction in Club Mahindra Resorts**

**Problem Statement**

Club Mahindra (Club M) makes significant revenue from Food and Beverages (F&B) sales in their resorts. The members of Club M are

Offered a wide variety of items in either buffet or À la carte form. Following are some benefits that the model to predict the spend by a

Member in their next visit to a resort will bring:

1. Predicting the F&B spend of a member in a resort would help in improving the pre-sales during resort booking through web and

Mobile app

1. Targeted campaigns to suit the member taste and preference of F&B
2. Providing members in the resort with a customized experience and offers
3. Help resort kitchen to plan the inventory and food quantity to be prepared in advance

Given the information related to resort, club member, reservation etc. the task is to predict average spend per room night on food and

Beverages for the each reservation in the test set.

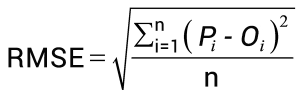
**Data Description**

* train.csv contains the training data with details on a set of reservations with the average spend per room night
* Data\_Dictionary.xlsx contains a brief description of each variable provided in the training and test set.

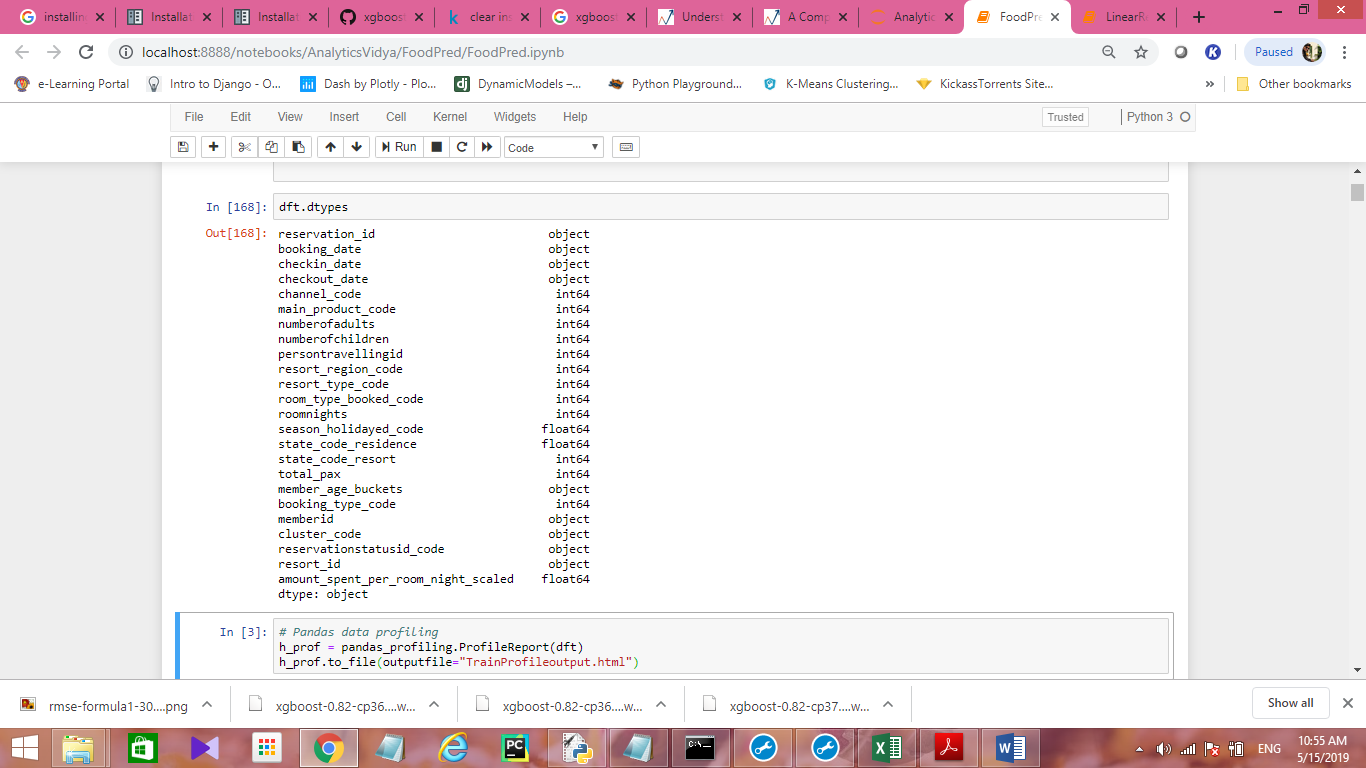
**Evaluation Metric**

Submissions are evaluated on 100 \* Root Mean Squared Error (RMSE) (https://www.analyticsvidhya.com/blog/2016/02/7-importantmodel-

evaluation-error-metrics/) on the variable amount\_spent\_per\_room\_night\_scaled.



**Data Analysis and Data Preparation**



**Numerical Attributes**

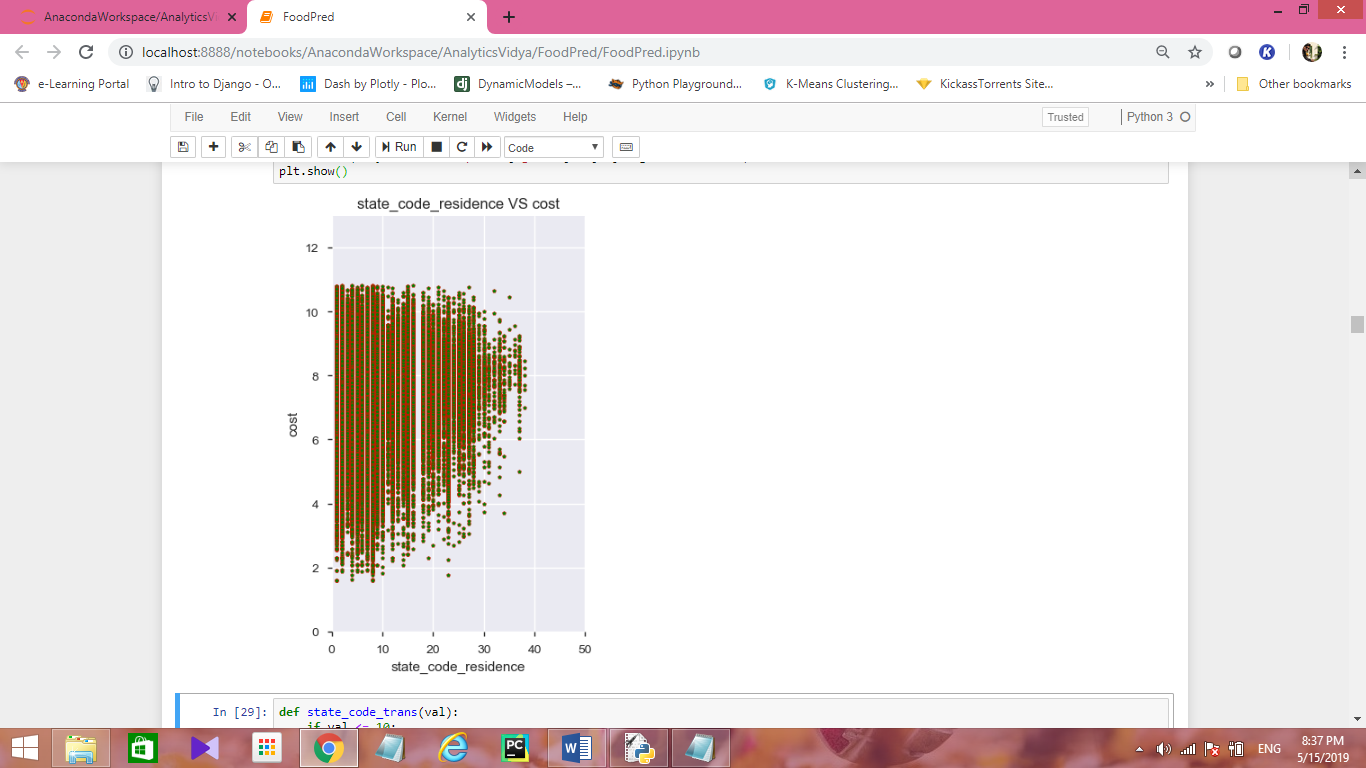
reservation\_id, numberofadults, numberofchildren, persontravellingid, roomnights, total\_pax, memberid, reservationstatusid\_code, resort\_id, amount\_spent\_per\_room\_night\_scaled.

**Categorical Attributes**

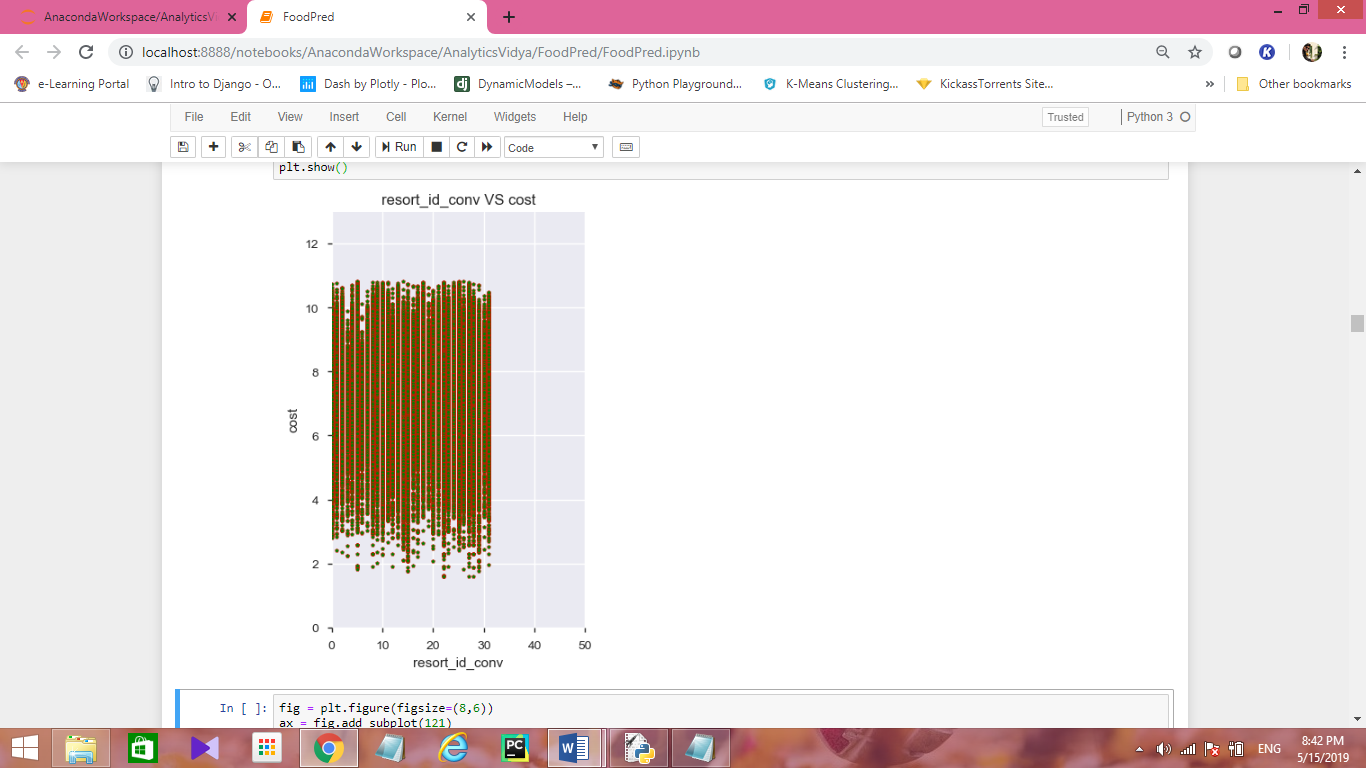
channel\_code, main\_product\_code, resort\_region\_code, resort\_type\_code, room\_type\_booked\_code, season\_holidayed\_code, state\_code\_residence, state\_code\_resort, booking\_type\_code, cluster\_code

**Target Variable:** amount\_spent\_per\_room\_night\_scaled

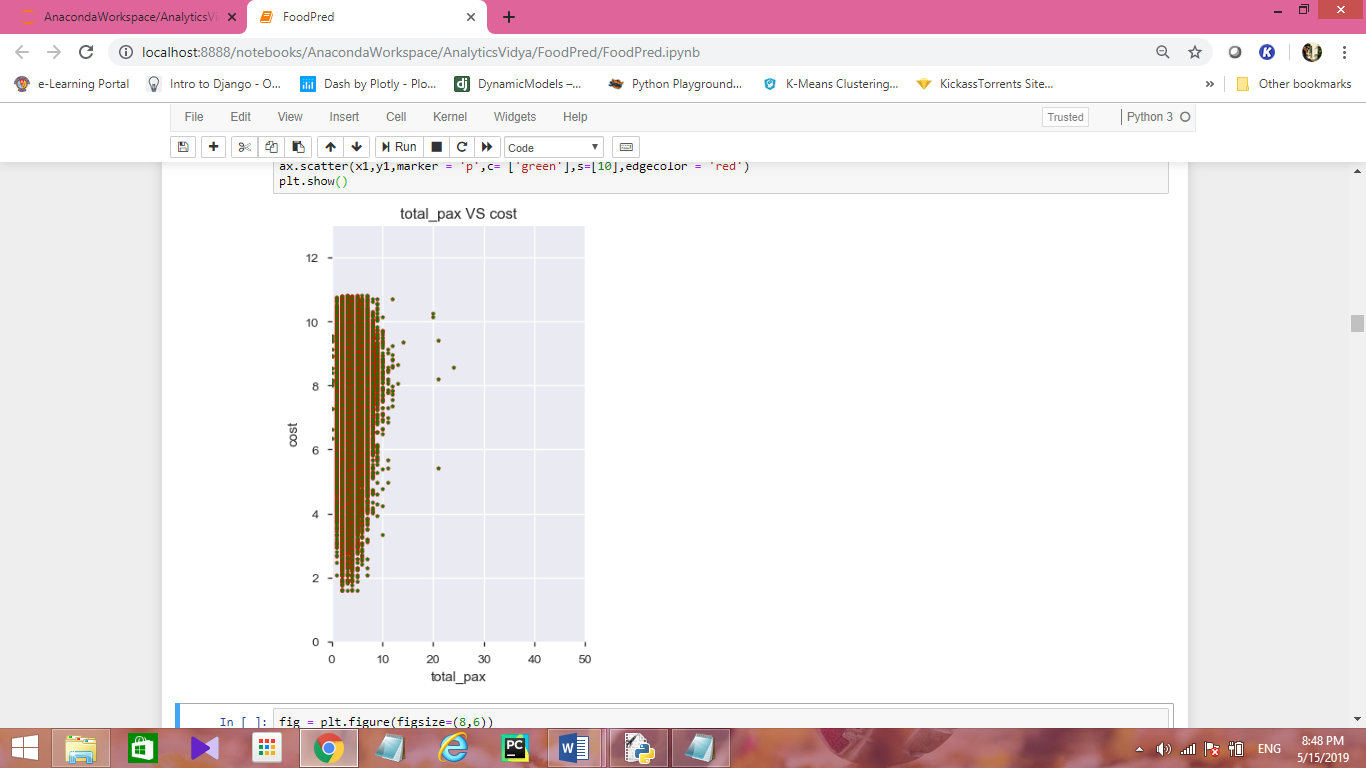
**Distribution of state\_code\_residence w.r.t amount\_spent\_per\_room\_night\_scaled**



**Distribution of Resort\_id\_conv w.r.t amount\_spent\_per\_room\_night\_scaled**



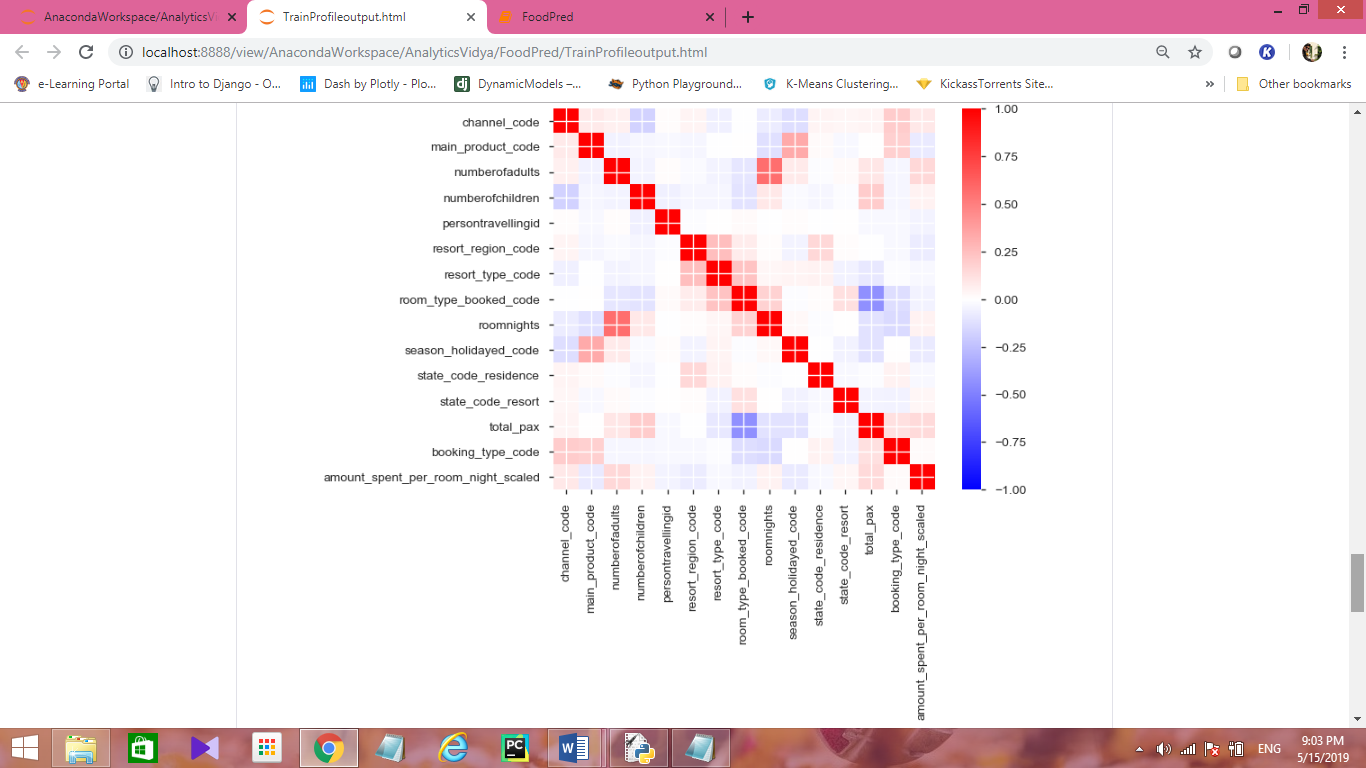
**Distribution of total\_pax w.r.to amount\_spent\_per\_room\_night\_scaled**



**Pearson Correlation**

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables. There are several types of correlation coefficient. Pearson correlation measures the linear association between continuous variables. In other words, this coefficient quantifies the degree to which a relationship between two variables.

* 1 indicates a strong positive relationship.
* -1 indicates a strong negative relationship.
* A result of zero indicates no relationship at all.



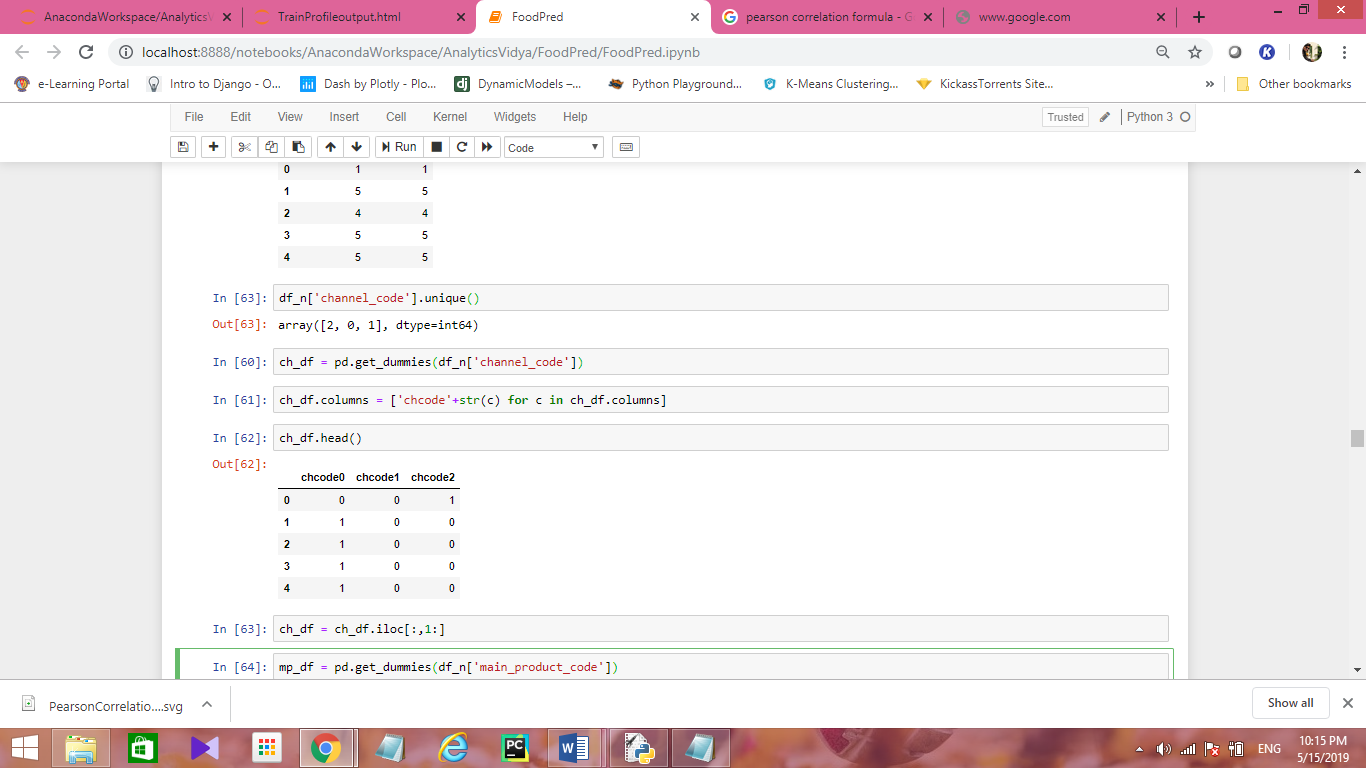
**Missing value features**

state\_code\_residence, state\_code\_resort, season\_holidayed\_code

**Creating Dummy Variables for Categorical Features**

Creation of dummy variables for categorical variables plays important part in regression techniqies.

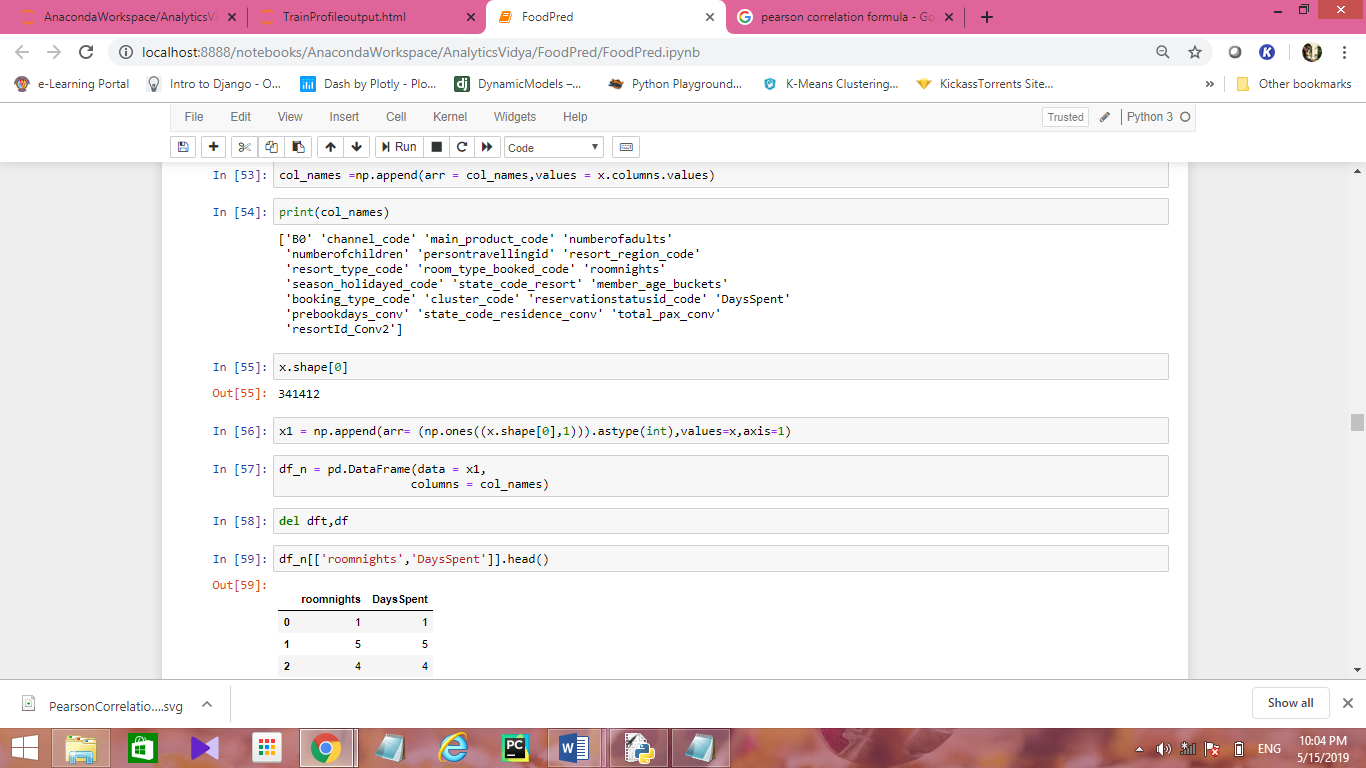
Some of the regression models can treat the categorical variables as ordinal feature and results in false predictions. To avoid the false predicitons we can create the dummy variables.



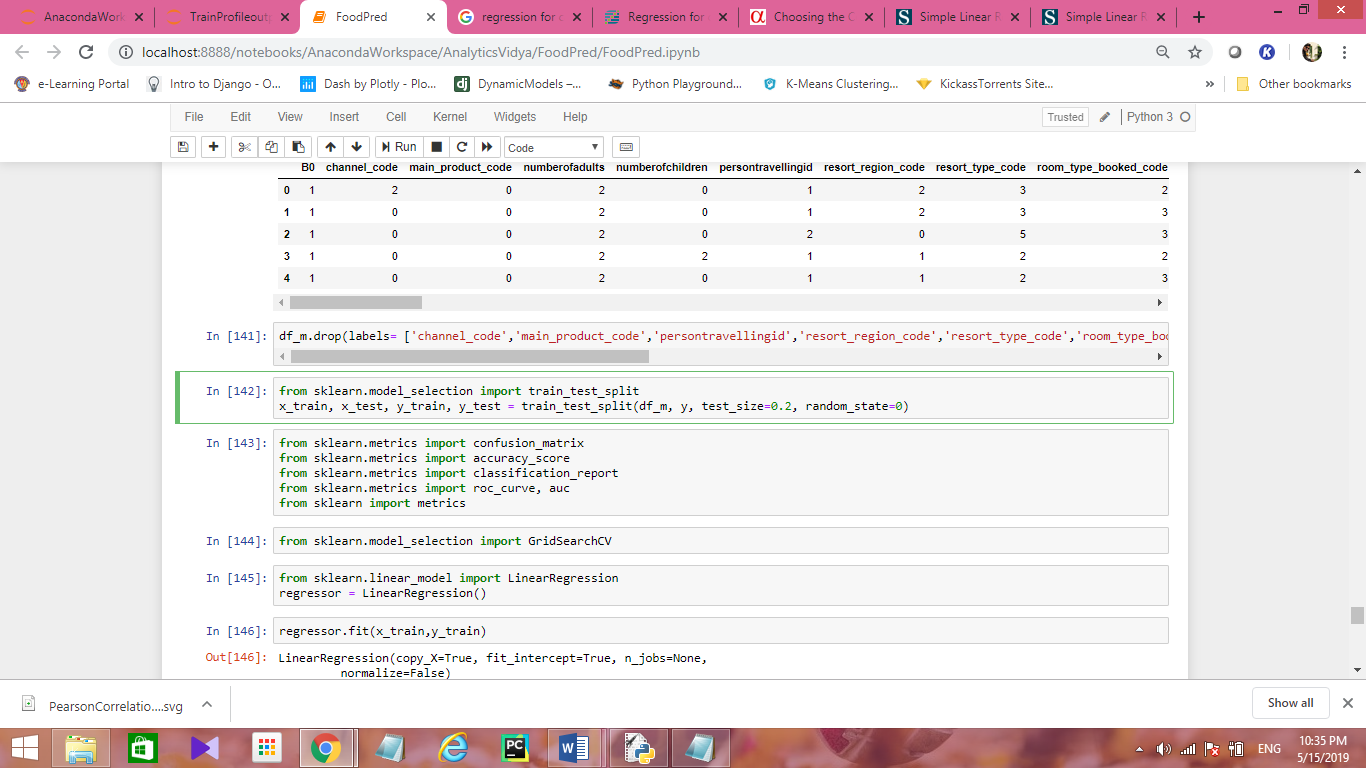
In the same way create the dummy variables for all categorical variables.

**Adding B0 (Intercept)Feature to data frame**

Y = B0+B1X1+B2X2+B3X3

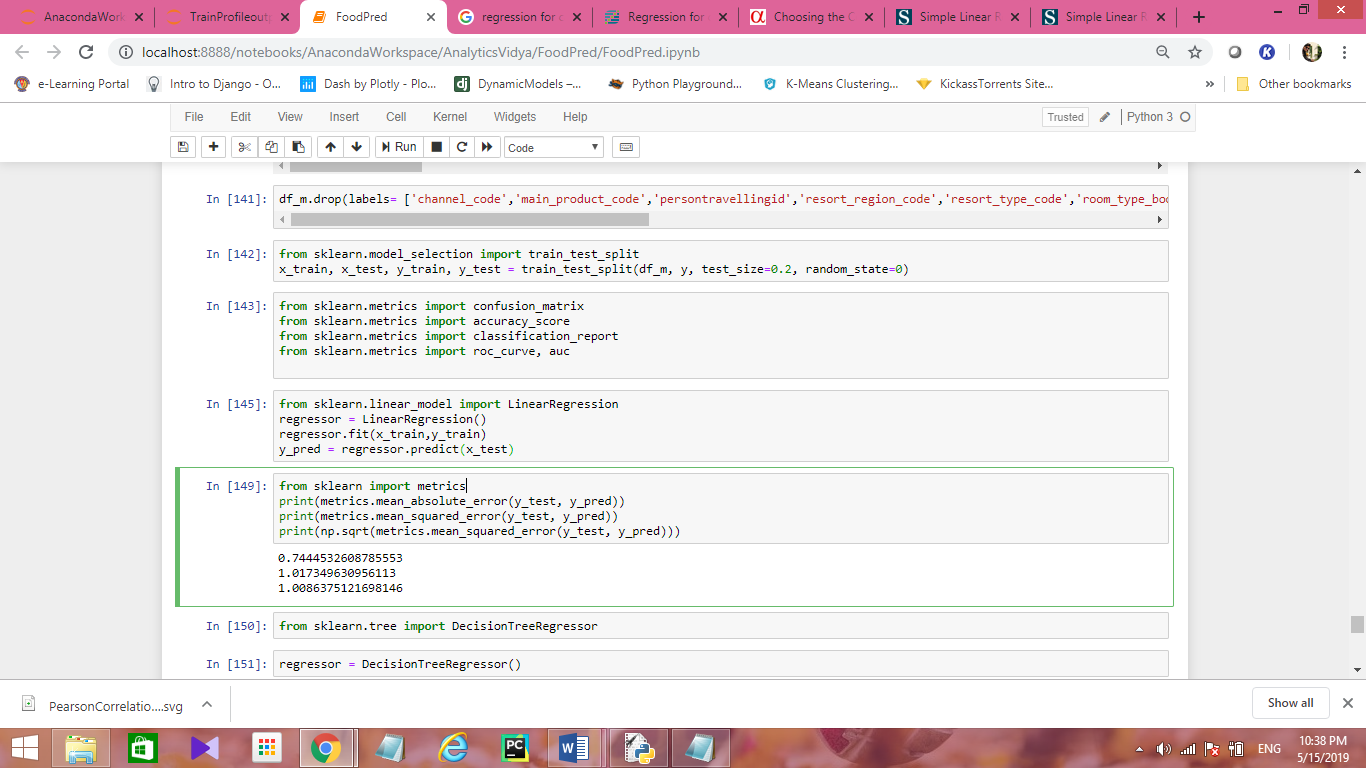


**Splitting the data set into train and test sets**



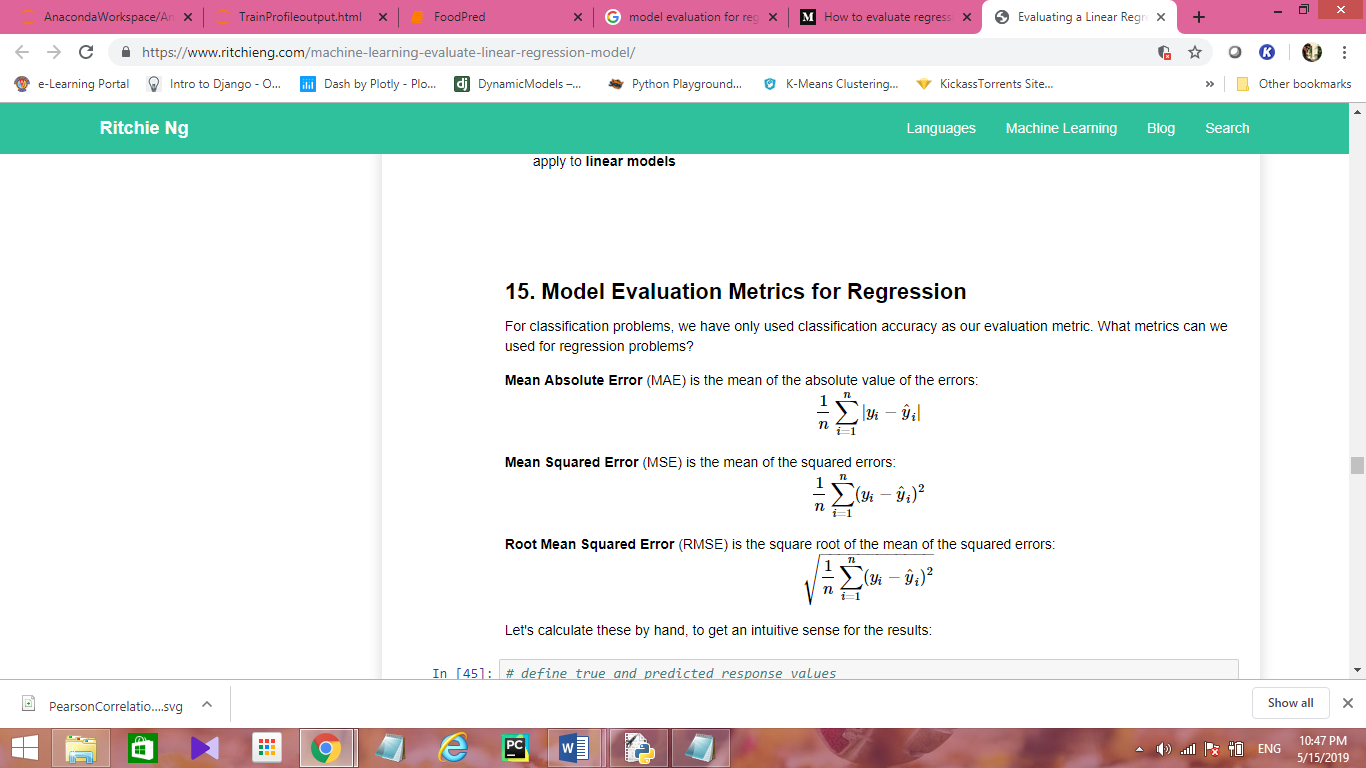
**Model Building**

**Selected Model:** LinearRegression (MultiLinearRegression)

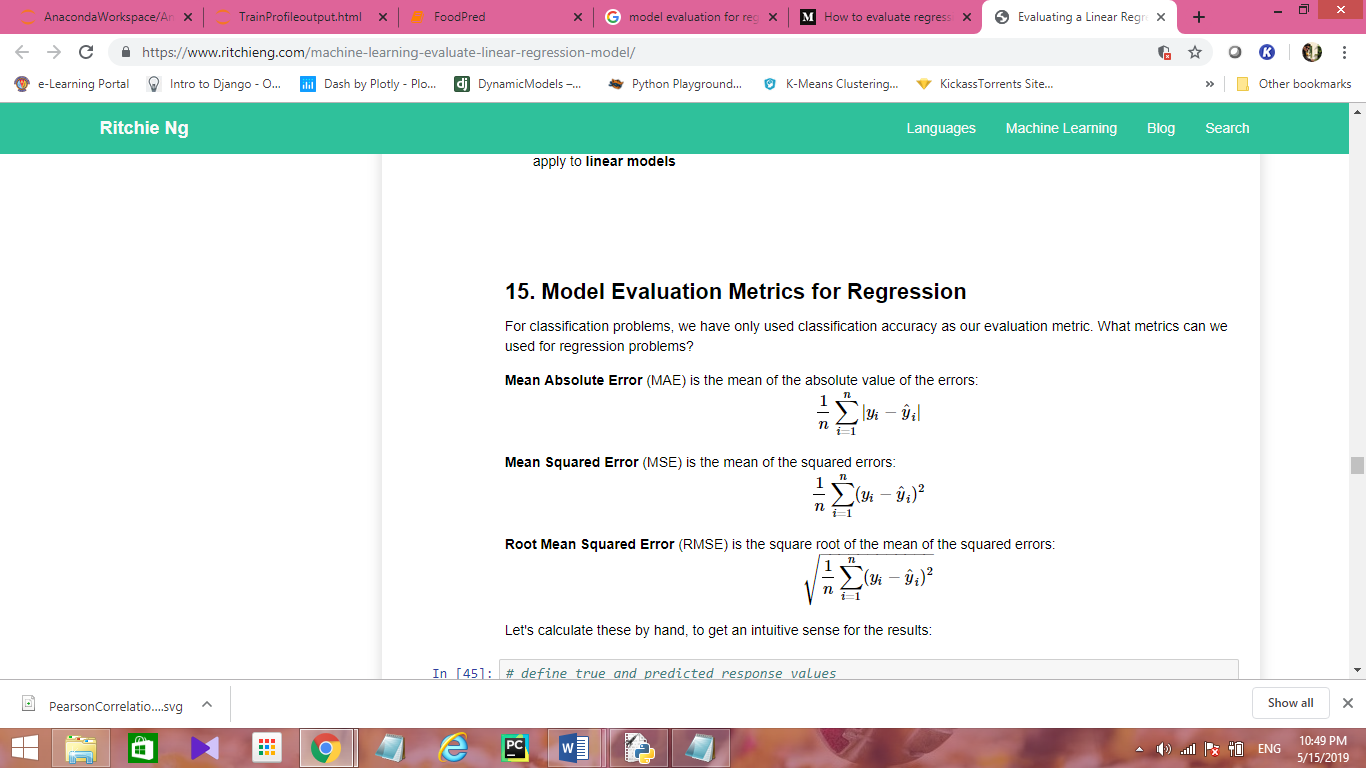


**Model Evaluation**

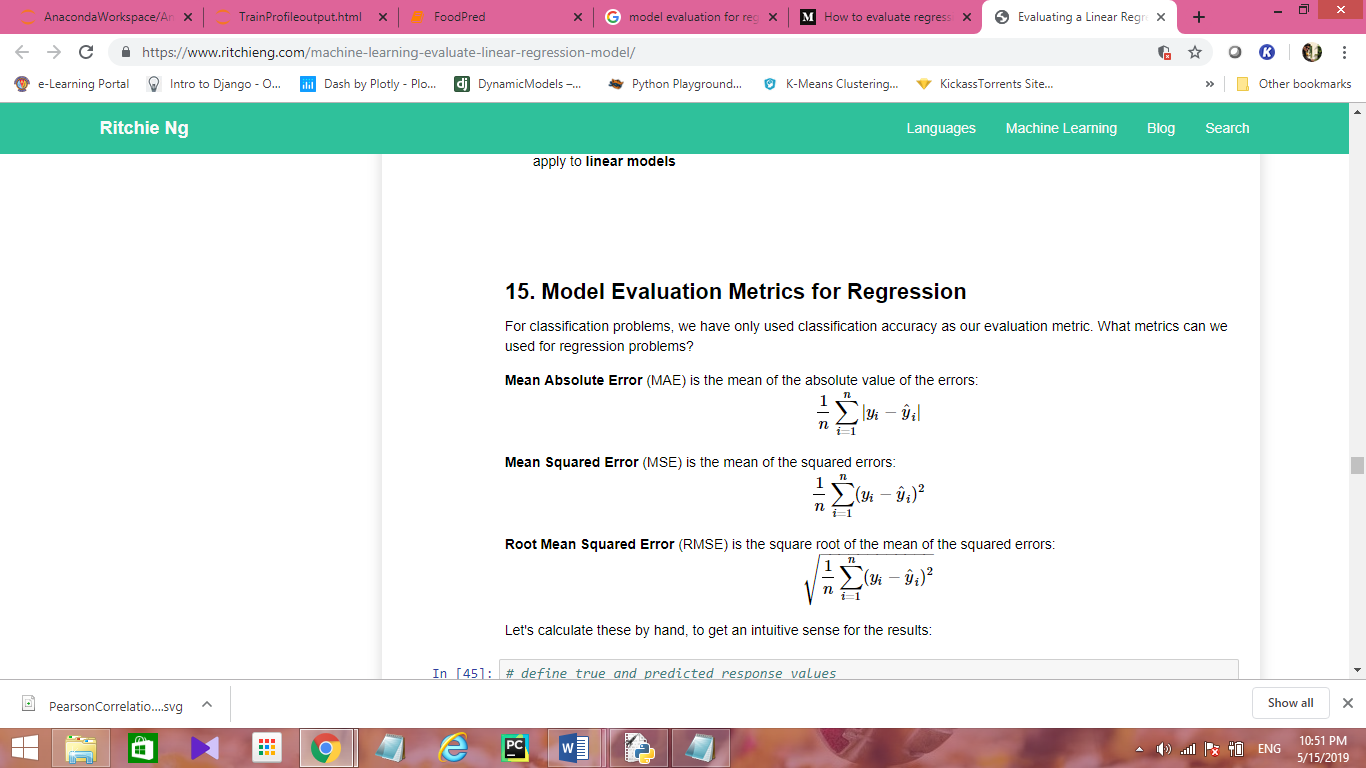
**Mean Absolute Error** (MAE) is the mean of the absolute value of the errors

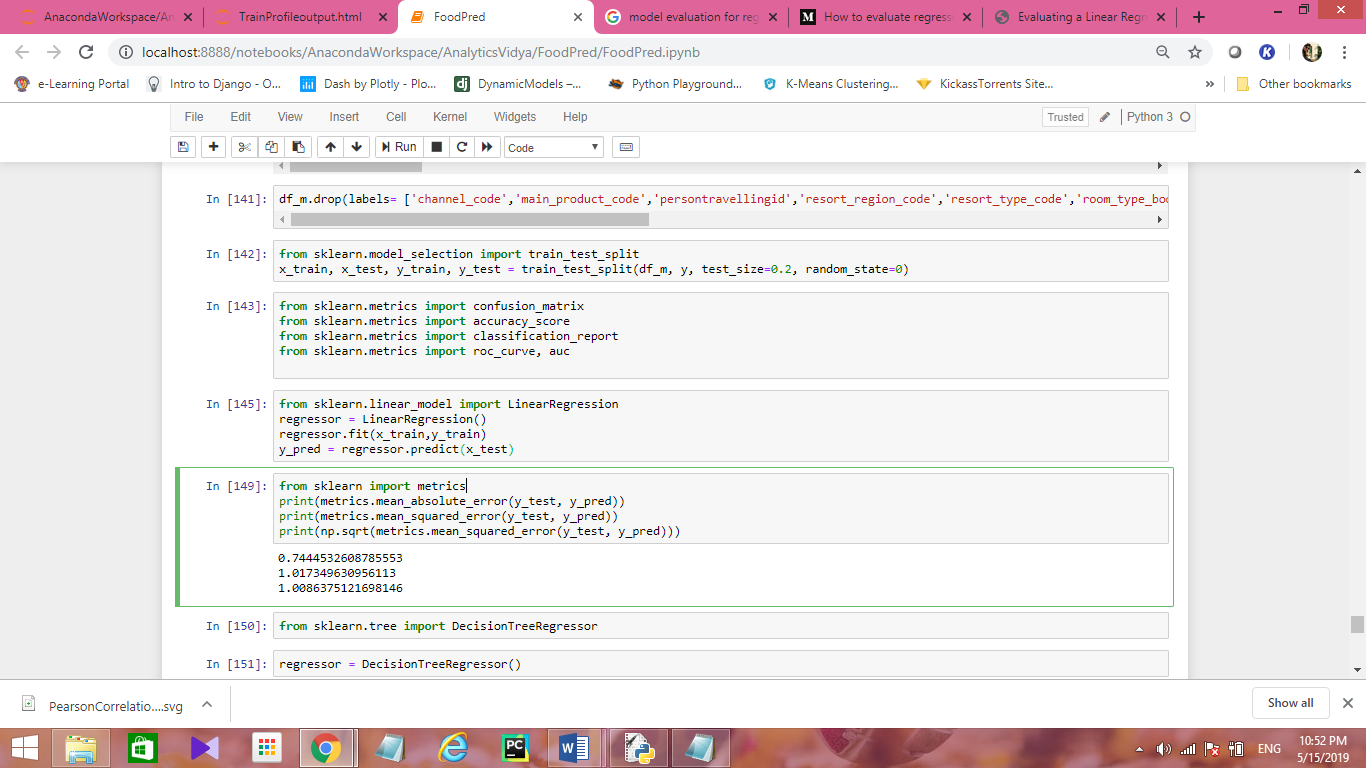


**Mean Squared Error** (MSE) is the mean of the squared errors



**Root Mean Squared Error** (RMSE) is the square root of the mean of the squared errors





We can observe that

MAE = 0.74

MSE = 1.01

RMSE = 1.00

It’s a good score, our model is able to predict the Food/ Beverages spent with less error! ☺