



Specialty Coffee Prediction

Bootcamp Data Science Project

Done by :
Rawan Mohammed
Ghaida Alamri

Introduction

Coffee is a beverage highly appreciated in all around the world due to its natural product with varied aromas and tastes.

We all have those friends who have a picky taste for coffee. They despise Starbucks, make their coffee on a scale, and drink specialty coffee. they assess beans farming, processing, roasting, and grinding based on specific standers. Well, I am one of those.

I've decided to combine my love of coffee with my passion for Machine Learning, and trying to find the answer to one of the worlds' most pressing question:

What makes that perfect cup of coffee?



Project Goal

We are going to build a model will be used to predict the total cup point value, where this value determines whether a coffee product can be considered “great” specialty coffee.

Tools :

Pandas, Matplotlib, Seaborn,
Sklearn



Dataset

The data set, originally uploaded from Coffee Quality Database, was re-posted to Kaggle

The data obtained is still in the form of raw data containing attributes such as company, mill, farm name, and the others.

The features used as input are the country of origin, processing method, moisture, and the plantation's altitude.





Data processing:

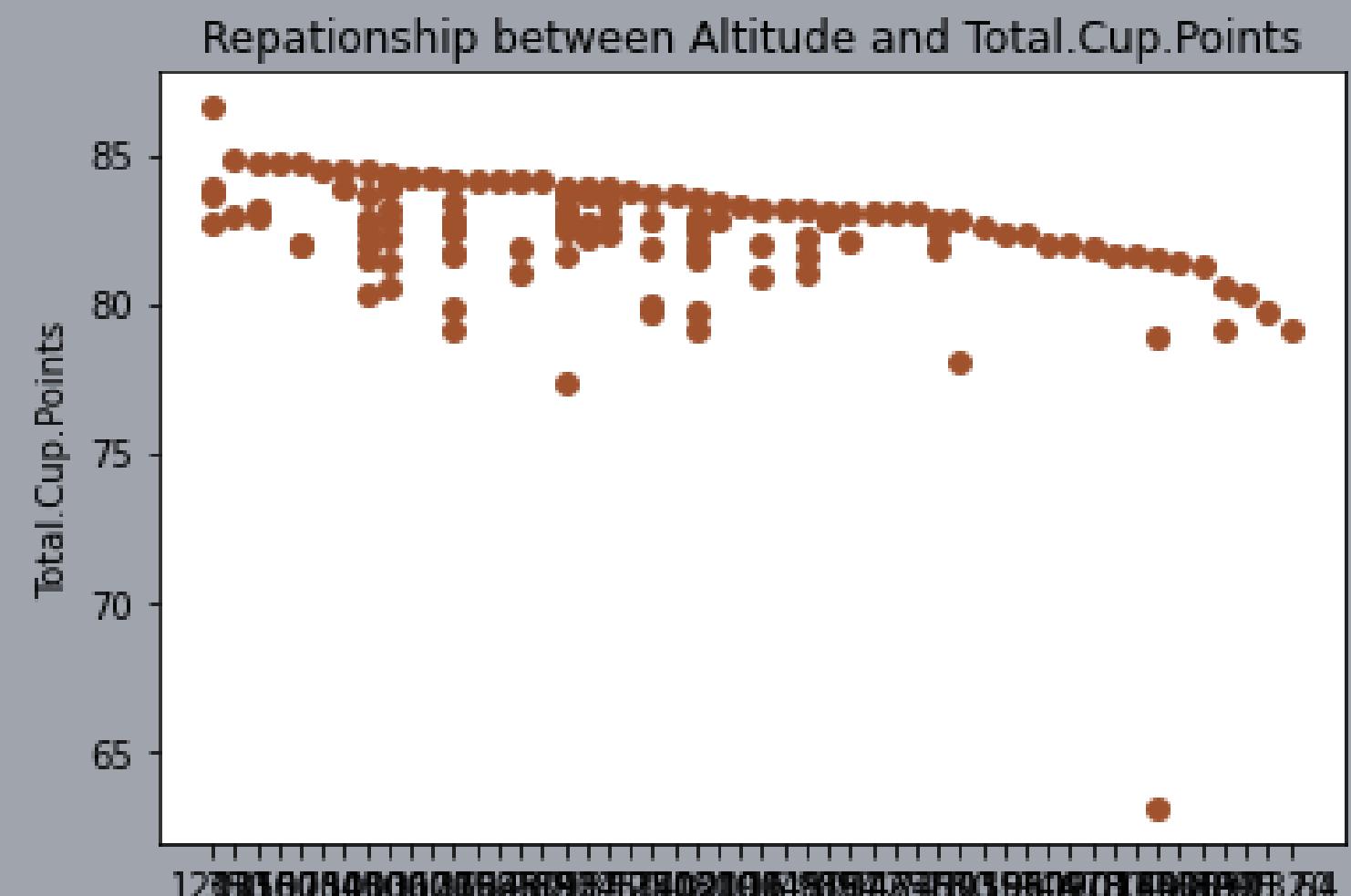
- Drop first column
 - Select target columns
 - Select important columns
 - Convert categorical data to numerical data.
-



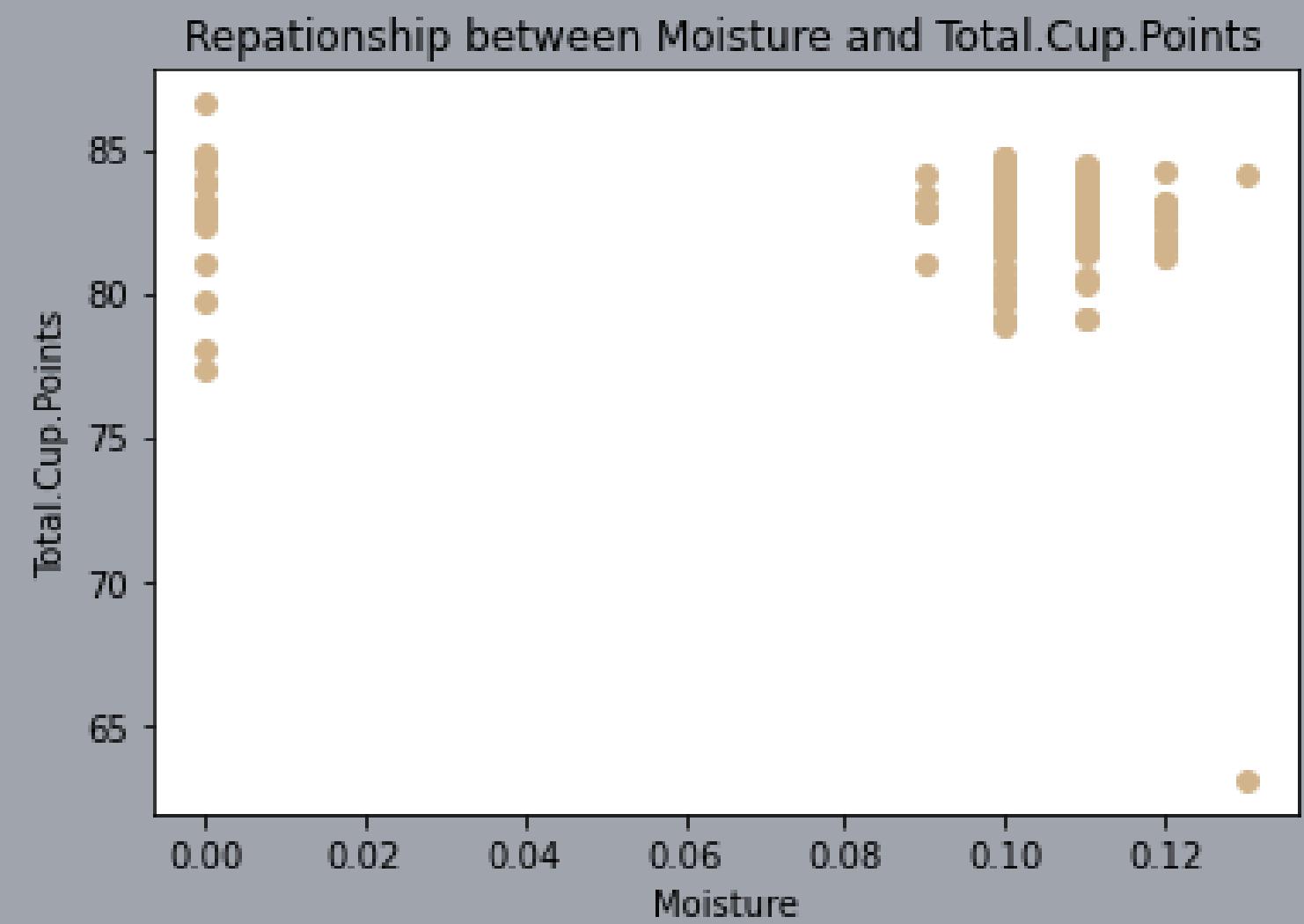
Can any of these features help predict total cup points ?
and what should be considered an exceptional coffee?

Data Visualization

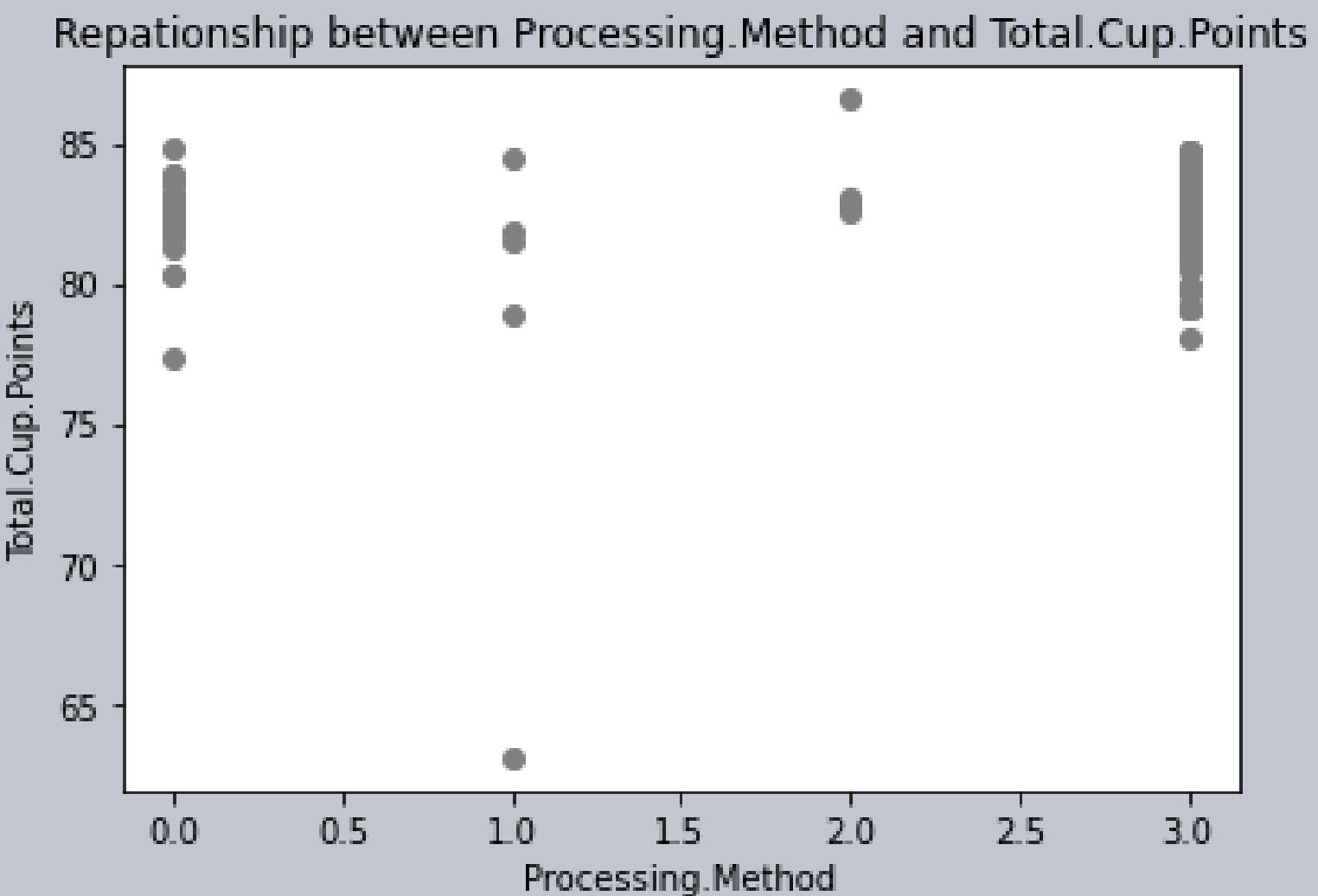
1- Relationship between Altitude and target:



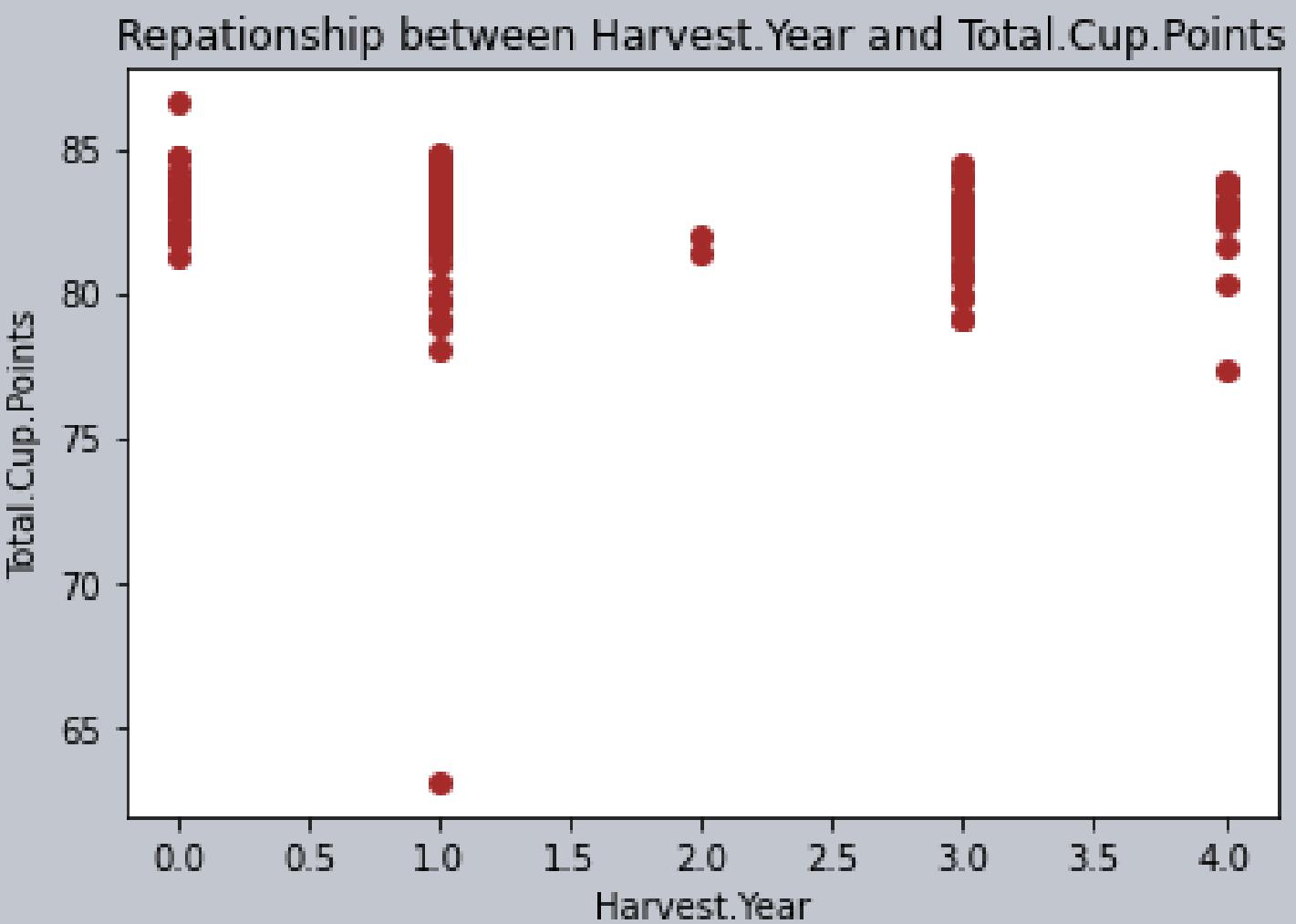
2- Relationship between Moisture and target:



3- Relationship between Processing method and target



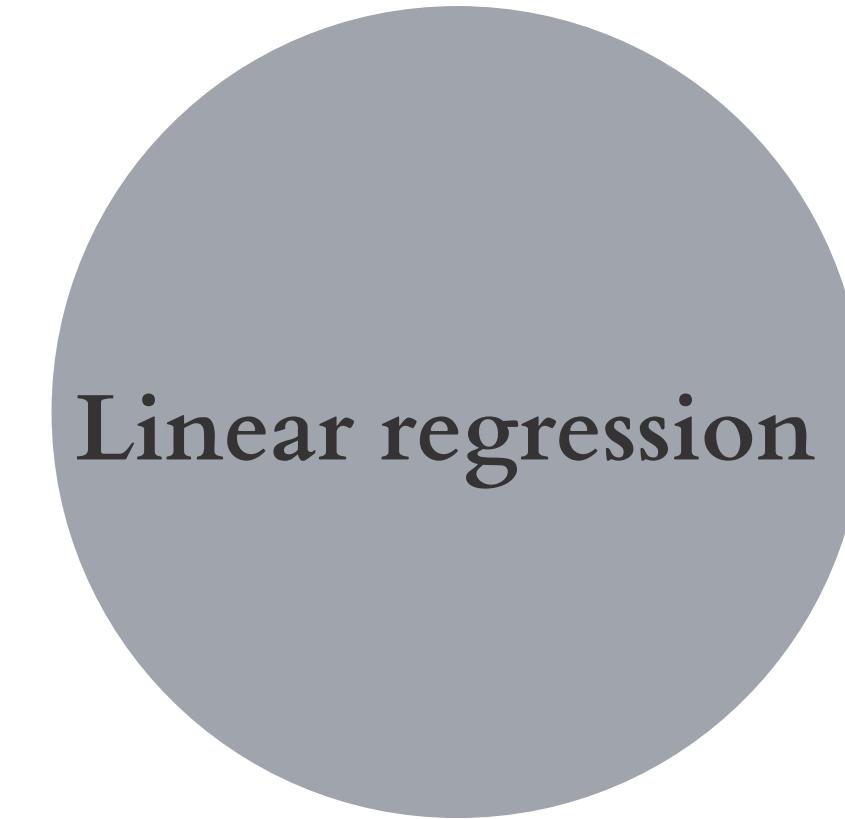
4- Relationship between Harvest year and target



I deemed any coffee that had a score higher than 80 to be a “great” Specialty coffee.



Machine Learning Model



Models Comparing:

SVM regression

Mean squared error for SVR regression: 3.27

Coefficient of determination for SVR regression: -0.03

linear regression

Mean squared error for linear regression: 15.73

Coefficient of determination for linear regression: -3.94





Conclusion

We see that SVM is better than linear regression because the mean square error and r2 score in SVM is less than it in Linear regression

So while the project does not look perfect , it was still fun to work on, and a great learning experience.

It seems like people's coffee preferences are just too varied to pick a clear winner, or do we just not have enough data?



Thank you