Project Report

Introduction:

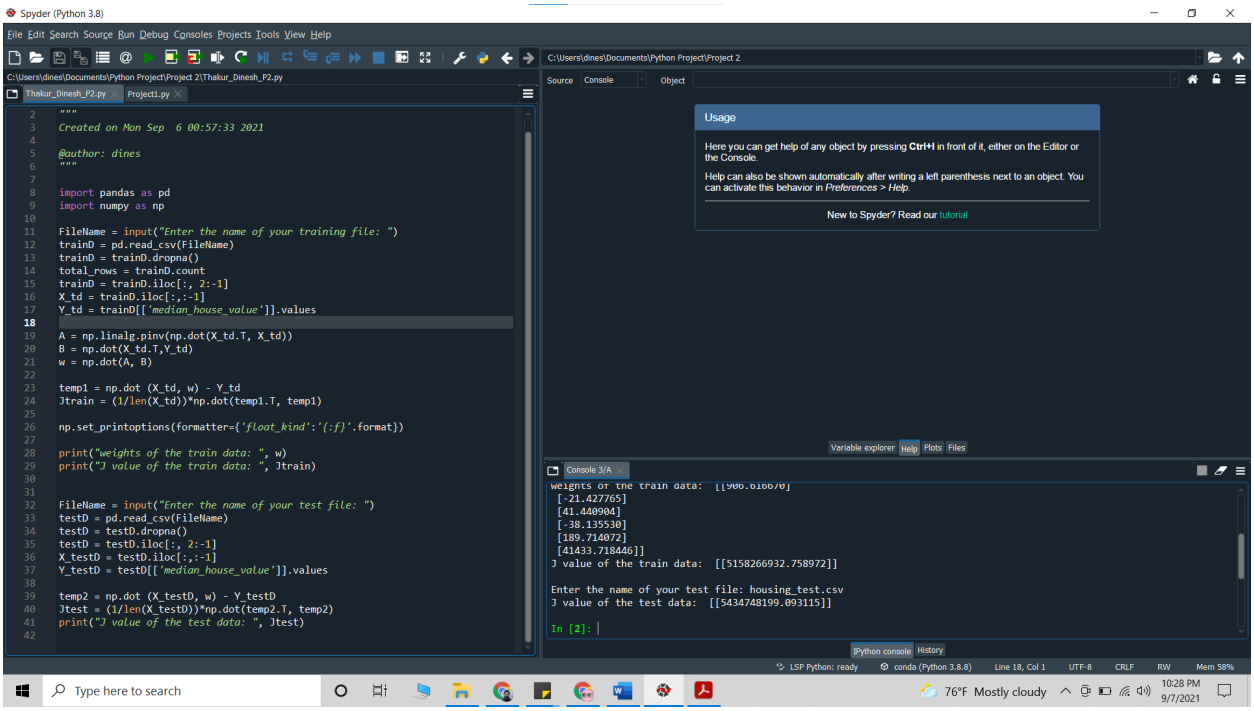
For both the training and testing dataset, I have taken the housing.csv file from Kaggle.com, which contains data of California Housing Prices for the following feature:

1. housingMedianAge: Median age of a house within a block; a lower number is a newer building
2. totalRooms: Total number of rooms within a block
3. totalBedrooms: Total number of bedrooms within a block
4. population: Total number of people residing within a block
5. households: Total number of families, a group of people living within a home unit, for a block
6. medianIncome: Median income for households within a block of houses (measured in tens of thousands of US Dollars)
7. medianHouseValue: Median house value for homes within a block (measured in US Dollars)

Independent Variable: housingMedianAge, totalRooms, totalBedrooms, population, households, medianIncome

Dependent Variable: medianHouseValue

I have split the first 5,000 rows and saved it as housing_train.csv to train my algorithm and find out weight and J value for programming purposes. After that, I have split the rows from 10,000 to 12,000 and saved it as housing_test.csv for testing my weights and deriving the new J value.



	Predicted output	Ground truth
0	146756.038270	201700
1	133753.477816	162500
2	180471.547838	209600
3	155537.464521	182100
4	188794.783584	206400
...	...	...
1979	138516.420512	153100
1980	155312.793904	75000
1981	167321.532010	203200

J value of the train data: [[5158266932.758972]]

J value of the test data: [[5434748199.093115]]