```
#! /usr/bin/python

import numpy as np
import pandas as pd

import statsmodels.api as sm
import statsmodels.formula.api as formula

import math

data = pd.read_csv("Auto.csv")

# Split dataset into halves.
train, test = np.split(data.sample(frac=1), [int(0.5*len(data))])

# Prepare model.
lm_train = formula.ols("mpg ~ horsepower", data=train).fit()
lm_test = formula.ols("mpg ~ horsepower", data=test).fit()

predictions = lm_train.predict()
mean_full = sum(predictions)/len(predictions)
print mean_full

poly1 = np.polyfit(train.loc[:, "horsepower"], train.loc[:, "mpg"], 2)
mean_poly1 = sum(poly1)/len(poly1)
print mean_poly1


poly2 = np.polyfit(train.loc[:, "horsepower"], train.loc[:, "mpg"], 3)
mean_poly2 = sum(poly2)/len(poly2)
print mean_poly2


```