Day1

1. Assignment
   Use the following link and answer the following questions.
   https://raw.githubusercontent.com/justmarkham/pandas-videos/master/data/chipotle.tsv
   1. What is total item price.
   2. What is minimum, maximum, average, 25% and 75% of item price
   3. Display all rows with item_name starts with “Chips”
2. Use tips dataset from seaborn, and display graph to show male and female smokers and non smokers
   a. Find daywise min, max, average, and 25 and 75 percentile of total bill
   b. On every data how many males and how many females go to hotel
   c. On which day number of visitors are maximum and on which day the number of visitors are min.

Day 3

1. Using following equation generate data for polynomial regression.
   X = 6 * np.random.rand(200, 1) - 3
   y = 0.7 * X**3 +0.8 * X**2 + 0.85 * X + 2 + np.random.randn(200, 1)

2. Find accuracy score for degree2,3, 4, 5,6,7,8 for above data, display best suitable degree for the scenario, also display coefficients.
3. Use **hiring.csv**. This file contains hiring statics for a firm such as experience of candidate, his written test score and personal interview score. Based on these 3 factors, HR will decide the salary. Given this data, you need to build a machine learning model for HR department that can help them decide salaries for future candidates. Using this predict salaries for following candidates,

   **2 yr experience, 9 test score, 6 interview score**

   **12 yr experience, 10 test score, 10 interview score**

   ## Answer
   53713.86 and 93747.79
4. Predict canada's per capita income in year 2020. There is an exercise folder here on github at same level as this notebook, download that and you will find canada_per_capita_income.csv file. Using this build a regression model and predict the per capita income fo canadian citizens in year 2020

## Answer
41288.69409442

**Day4:**

1. **Use hiring.csv and home prices apply DecisionTreeRegressor, SupportVectorRegressor and Random forest and find the accuracy score**

**Day 5:**

1. **Using Iris data set, apply Naïve bays theorem and KNN algorithm, and find accuracy of the model**
2. **From Iris dataset keep rows with values Setosa-virginica and setosa-bersicolor, and apply logistic-regression algorithm and find accuracy score.**

**Day 6 :**

**Use wine dataset, and cancer dataset apply all classification models,**
**Use KFold cross validation with value of k=10, and display the conclusion which model is better.**

**Also Find what should be the optimum value of n_estimator in RandomForestClassifier, (use KfoldCross validation)**

**In the assignment for every step, add the explanation.**

Day 7:
Use Cancer.csv and Iris dataset apply all Classification models with AdaBoosting Algorithm, Find which is best suitable.

Day 8:

Explore the XGBoost algorithm in python using the sci-kit learn package. For that reason, we would take the help of a dataset from the UCI Machine Learning repository. It's called "Pima Indians Diabetes Database ". This dataset is from the National Institute of Diabetes, India. The objective of our XGBoost model would be to predict whether or not a patient has diabetes, based on certain diagnostic measurements such as BMI, insulin level, age, skin, blood pressure, and so on. The dependent variable is a 0/1 binary flag, where 0 stands for a non-diabetic patient and 1 means the patient has diabetes.
More details on the data: https://www.kaggle.com/datasets/uciml/pima-indians-diabetes-database