

Vodafone Final Exam

Before you start, please download “**auto-mpg.data.txt**” dataset from SUCourse;
Dataset has following columns and information.

mpg	cylinders	displacement	horsepower	weight	acceleration	model_year	origin	name
18	8	307	130	3504	12	70	1	chevrolet chevelle malibu
15	8	350	165	3693	11.5	70	1	buick skylark 320
18	8	318	150	3436	11	70	1	plymouth satellite
16	8	304	150	3433	12	70	1	amc rebel sst

You should create a ipynb file for each question and upload them to SUCourse at the end of the exam.
(Please do not use Turkish characters in file names)

Your_name_lastname_q1.ipynb

Your_name_lastname_q2.ipynb

Your_name_lastname_q3.ipynb

1. Find the cars with the worst fuel efficiency (lowest mpg) for each **origin**. (10 points)
(1→USA, 2→Europe, 3→Asia)
2. Add a new column named “Car-Type” that has following values according to acceleration.
(User Defined Function) (10 points)
(0 - 7 secs → Fast Car, 7 - 12 secs Average Car, 12+ secs Slow Car)
3. We want to predict **mpg** for given automobile info. (40 points)
Choose one of the ML algorithms from Spark ML library and prepare data for training.
 - a. Please explain each step with a comment before each step.
(E.g. : Dropping column because ... , Replacing null values with ... etc.)
 - b. **origin** column should be one hot encoded
 - c. **mpg** column is the label value.

Create a model and print your test accuracy or RMSE. (Evaluation)