**Overview**

With the rise in different career opportunities, it's a challenging balance between work and personal life. We are learning and upskilling ourselves to grow ourselves in the competitive world. The wages are linked to the working hours and the skills we learn over time. The working hours vary at different locations with respect to earn the same income.

**Problem Statement**

In this hackathon we need to  predict the working hours per week at different locations with attributes such as workclass, education, marital-status, occupation capital-gain, capital-gain, capital-loss etc. to get the desired salary in a range.

**Data:**

* Age: in years
* Workclass: Private, Self-emp-not-inc, State-gov, Self-emp-inc, Federal-gov, Local-gov, Never-worked, Without-pay
* Fnlwgt:  number of people the census believes
* Education: HS-grad, Bachelors, Masters, Some-college, 7th-8th, Doctorate, Prof-school,  Assoc-voc, Assoc-acdm, 5th-6th, 11th, 9th, 1st-4th, 12th, 10th, Preschool
* Education-num:  represents a numerical value of education
* Marital-status: Never-married, Married-civ-spouse, Divorced, Married-spouse-absent, Separated, Widowed, Married-AF-spouse
* Occupation: Other-service, Sales, Transport-moving, Prof-specialty, Exec-managerial, Adm-clerical, Handlers-cleaners, Farming-fishing, Craft-repair, Protective-serv, Machine-op-inspct, Tech-support, Priv-house-serv, Armed-Forces
* Relationship: Unmarried, Husband, Not-in-family, Own-child, Wife, Other-relative
* Race: White, Amer-Indian-Eskimo, Asian-Pac-Islander, Black, Other
* Sex:  Female, Male.
* Capital-gain: negative or positive
* Capital-loss: negative or positive
* **Hours-per-week: in Hours (Target)**
* Native-country: United-States, Cuba, Philippines, Mexico, Honduras, Jamaica, England, India, Canada, Puerto-Rico, South, Guatemala, Haiti Iran, Outlying-US(Guam-USVI-etc), El-Salvador, Italy, Cambodia, Germany, Portugal, Japan, Yugoslavia, Ireland, China, Dominican-Republic, Taiwan, Ecuador, Poland, Trinadad&Tobago, Vietnam, Thailand, Columbia, Peru, France, Greece, Hungary, Laos, Scotland, Hong, Nicaragua,  Holand-Netherlands

**Train: 18944 rows x 15 columns**

**Test: 8119 rows x 14 columns**

**Target Column: “hours-per-week”**

**Skills:**

* *Optimize RMSE*
* *Forecasting*
* *Timeseries*
* *Machine Learning Approach*

**Evaluation**

**What is the Metric In this competition? How is the Leaderboard Calculated ?**

* The submission will be evaluated using the RMSE metric. One can use ‘np. sqrt ([Mean Squared Error)](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.mean_squared_error.html)’ to calculate the same. [mean\_squared\_error(y\_true, y\_pred, squared=False)]
* This hackathon supports private and public leaderboards
* This **hackathon** supports **private** and **public** leaderboards
* The **public leaderboard** is evaluated on **30%** of **Test data**
* The **private leaderboard** will be made available at the **end** of the **hackathon** which will be **evaluated** on **100% Test data**

**How to Generate a valid Submission File**

You should submit a .csv/.xlsx file with exactly 8119 rows with 1columns ***('hours-per-week***”***')***. Your submission will return an **Invalid Score** *if you have extra columns or rows*.

The submission limit for this hackathon is one account per participant.

**Note: Do not shuffle the sequence of the test.csv**

Using Pandas:

submission\_df.to\_csv('my\_submission\_file.csv', index=False)