Petavue

Hello,
Thanks for considering taking up an opportunity to be part of the Petavue family. We deeply appreciate your interest. Through the below assignment we hope to find the appropriate role for you at Petavue.

Assessment for the role: AI Research Intern at Petavue

Problem Statement

Extract the flight and weather data and merge both to create a file '*merged_data.csv*' which will have all the flight details and the weather data associated with it.

The weather folder consists of the weather data at 15 different airports in the USA between 2016 and 2017. The flight data of all flights inside the USA is provided. Merge these two data appropriately such that each record of the flight should have the corresponding weather data available.

Preprocess the data for the airports mentioned below for the years 2016 and 2017 alone

Airport codes:

ATL	CLT	DEN	DFW	EWR
IAH	JFK	LAS	LAX	MCO
MIA	ORD	PHX	SEA	SFO

Destination and origin of the flights should be only from the above airports.

Recommended Weather Columns:

WindSpeedKmph	WindDirDegree	WeatherCode	precipMM
Visibilty	Pressure	Cloudcover	DewPointF
WindGustKmph	tempF	WindChillF	Humidity
date	time	airport	

Recommended Flight Columns:

FlightDate	Quarter	Year	Month
DayofMonth	DepTime	DepDel15	CRSDepTime
DepDelayMinutes	Origin	Dest	ArrTime
CRSArrTime	ArrDel15 (label)	ArrDelayMinutes (target)	

Link to assets

Link to flight data: [here]
Link to weather data: [here]
Workbooks for NumPy, pandas and JSON: [here]

Time Limit: 6 hrs

Bonus:

Create a classifier that classifies whether a given flight will be delayed or not using the features. Use any two models from : Decision trees, Extra Trees Classifier, XGBoost, Random Forest. Additionally, contrast the 2 classification models' performances using metrics (of your choice) to evaluate the classification in ML

Communication

You are encouraged to contact your Petavue contact to seek any clarifications on any matter, before you begin your assignment or while completing your assignment. It is OK to ask questions if you don't understand something or if you get stuck. We encourage collaboration within the team and strongly believe that we all need to push and help each other every day to reach our goals. But it is also important that you have the ability to figure things out by yourself and don't need too much spoon-feeding. Striking the right balance is key.

All the best and we look forward to welcoming you to our team!!

FAQ

1. *Where do I upload my code?*
   Kindly upload the completed assignment to your personal Github within the stipulated time and send us the link to the same
2. *What is the expected result?*
   Post the top 5 rows and the shape of the final dataframe in your Readme.md file
3. *Should the bonus section be completed within the stipulated time*?
   Not necessary, completing the original problem statement is the ideal scenario. You can have a shot the bonus section if you have time to spare.