When tomorrow is the last date of assignment submission

UNIVERSITYSTUDENT.org

upGrad
#LifeKoKaroLift

Data Science Certification Program

upGrad

Today's Agenda

1. Assignment Walkthrough
2. Step by Step Approach
3. Doubt Session (QnA)

Linear Regression:

Bike-sharing system

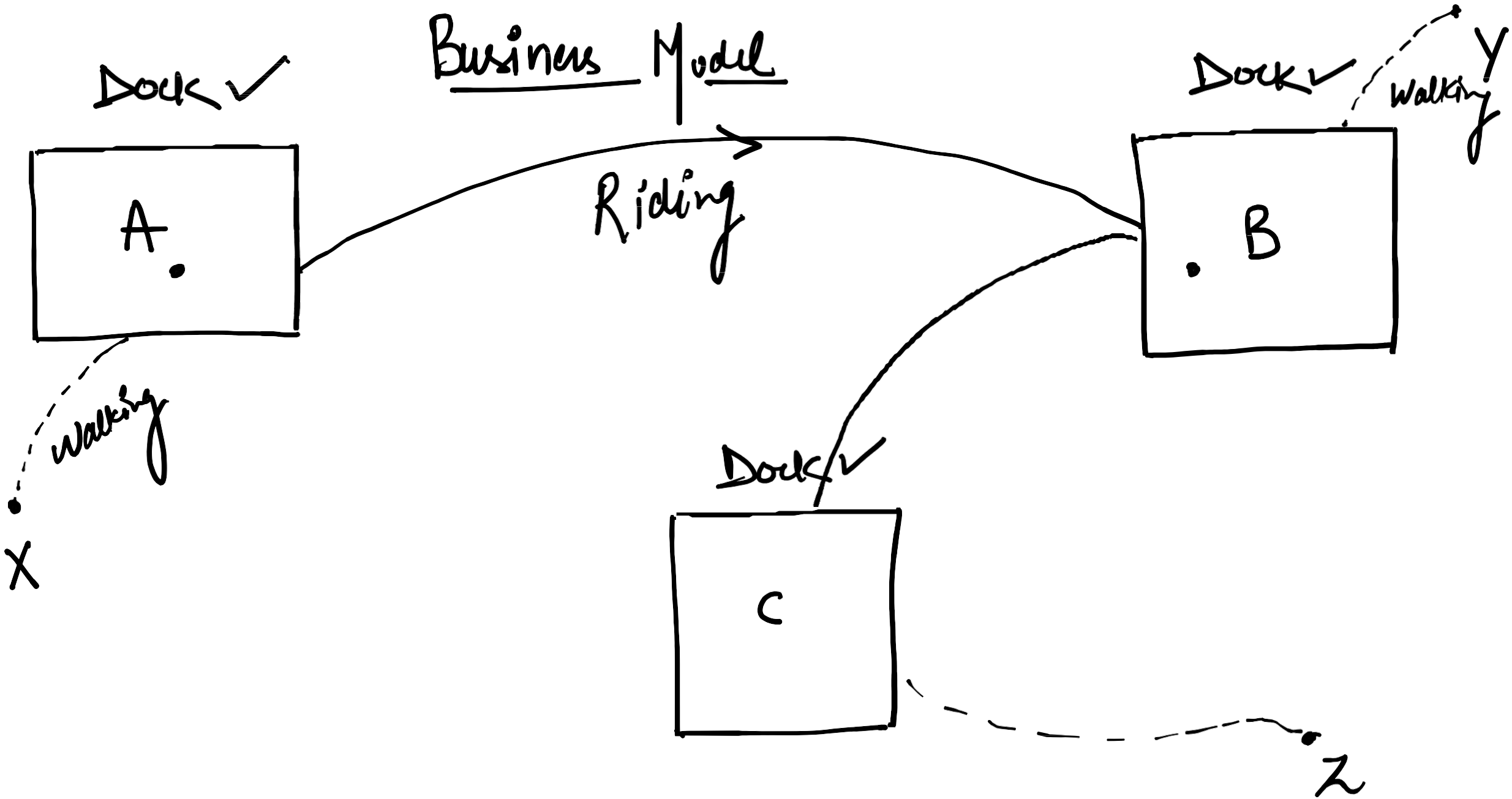
upGrad

Assignment Problem Statement

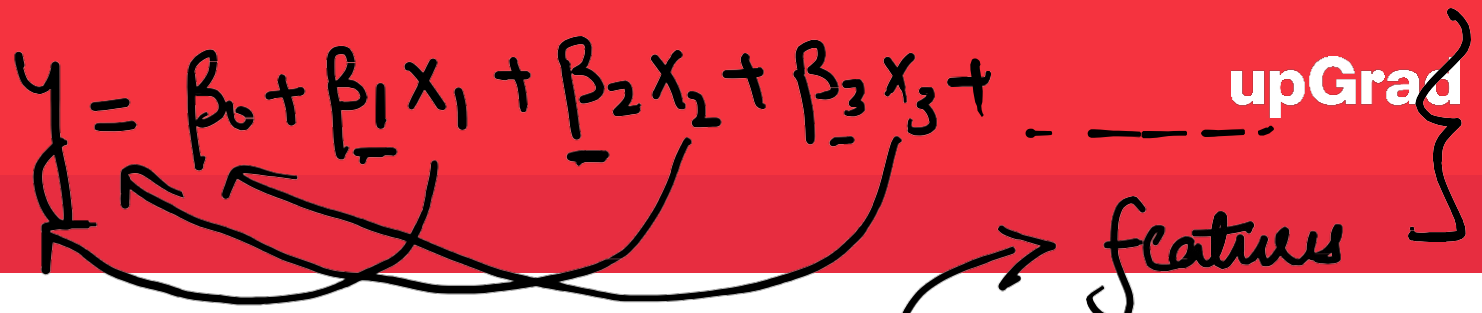
- A bike-sharing system is a service in which bikes are made available for shared use to individuals on a short-term basis for a price or free. Many bike share systems allow people to borrow a bike from a "dock" which is usually computer-controlled wherein the user enters the payment information, and the system unlocks it. This bike can then be returned to another dock belonging to the same system.

- A US bike-sharing provider BoomBikes has recently suffered considerable dips in their revenues due to the ongoing Corona pandemic. The company is finding it very difficult to sustain in the current market scenario. So, it has decided to come up with a mindful business plan to be able to accelerate its revenue as soon as the ongoing lockdown comes to an end, and the economy restores to a healthy state.

→ Yulu (Indian)


Business Model
Dock
A
Riding
Dock
B
Walking
Y
Walking
X
Dock
C
Z

Linear Regression:



- They have contracted a consulting company to understand the factors on which the demand for these shared bikes depends. Specifically, they want to understand the factors affecting the demand for these shared bikes in the American market. The company wants to know:

1. Which variables are significant in predicting the demand for shared bikes. → 1
2. How well those variables describe the bike demands → 2

What you need to do?

- Create a linear model that describe the effect of various features on price.
- The model should be interpretable so that the management can understand it.

} Doable

Target (y) → demand

Industry to decide which algorithm to use

Machine learning

Supervised Machine learning (Target variable is available)

Linear Regression (Target variable is continuous)

(Demand)

→ Expected Tasks :- (Step by Step approach)

1- Data Understanding and data loading :-

* day.csv → Data for assignment [pd.read_csv(path)]
* html link → Data Dictionary

2- Pre processing Steps :-

* Pls drop two columns Casual & Registerd } (Not features)
* Map all values of categorical variables from data dictionary.

Season	→	Season
1		Spring
1		Spring
2		Summer
3		Autumn
2		Summer
4		winter

{1: Spring, 2: Summer
3: Autumn, 4: Winter}

* Create the dummy variables for all categorical variables.
  where no. of categories > 2.

pd.get_dummies ( )

drop_first = True

dummy variables

Season

1 - Spring
2 - Summer
1 - Spring
3 - Winter
2
3

Spring	Summer	Winter
1	0	0
0	1	0
1	0	0
0	0	1
0	1	0
0	0	1

No. of dummy = n-1

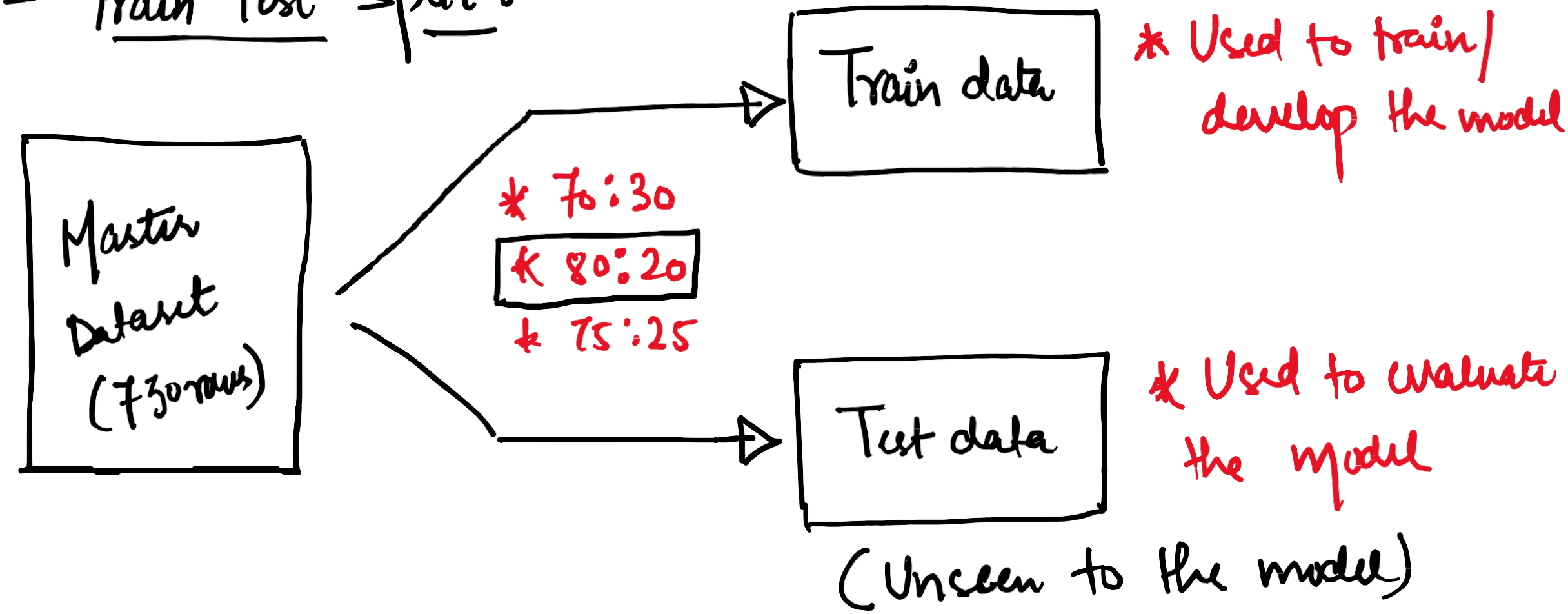
No. of Categories = 3

No. of dummies = 2

3- EDA :- (Mandatory & inclusive Step in data modelling)

* Univariate Analysis } → Mandatory
* Bivariate Analysis
* Multivariate Analysis } → optional

4- Train Test Split :-



5- Missing value Imputation (if any) (Can be skipped for this assignment)

6- Scaling (Mandatory for linear models)

- fit_transform (Traindata)
- transform (Test data)

7- Feature Selection :-

- Automated (RFE)
- Manual (p-value + VIF)
  - one by one variable
- Hybrid (Automatic + Manual)

Both of them

3 of A
1 5 f
M
1 0 f

→ Cutoff of p-value is 0.05 (p-value < 0.05)

→ Cutoff of VIF is 5 (VIF ≤ 5)

8- Model Building

9- Evaluation (Test data)

* R^2
* Adjusted R^2

} 80-85% (for train & test both)

Linear Regression:

upGrad

Data Preparation:

- Load the data and understand it using dictionary provided.
- Convert the columns to proper data types.
- Create dummies for categorical variables.

Model Building

- Divide the data to train and test.
- Perform scaling.
- Divide the data into X and y.
- Perform Linear Regression.
- Use mixed approach if you want.

Linear Regression:

upGrad

How to go about selecting features for a good model?

- RFE
- Manual
- Mixed

Assignment Steps

Model Evaluation

- Check the various assumptions.
- Check the Adjusted R-Square for both test and train data.
- Report the final model.

} final evaluation steps

Linear Regression: Subjective Questions

upGrad

Assignment-Subjective

Steps to answer subjective part

MS Word

- Answer all the questions.
- You can write the answer using any software but submit the file in PDF format .
- You can use images and plots to support your answer.
- Make sure the question is answered with sufficient number of word: No limit
- Please don't copy for any online available literature.

Linear Regression:

(Jupyter Notebook + Subjective) (PDF)
.ipynb file

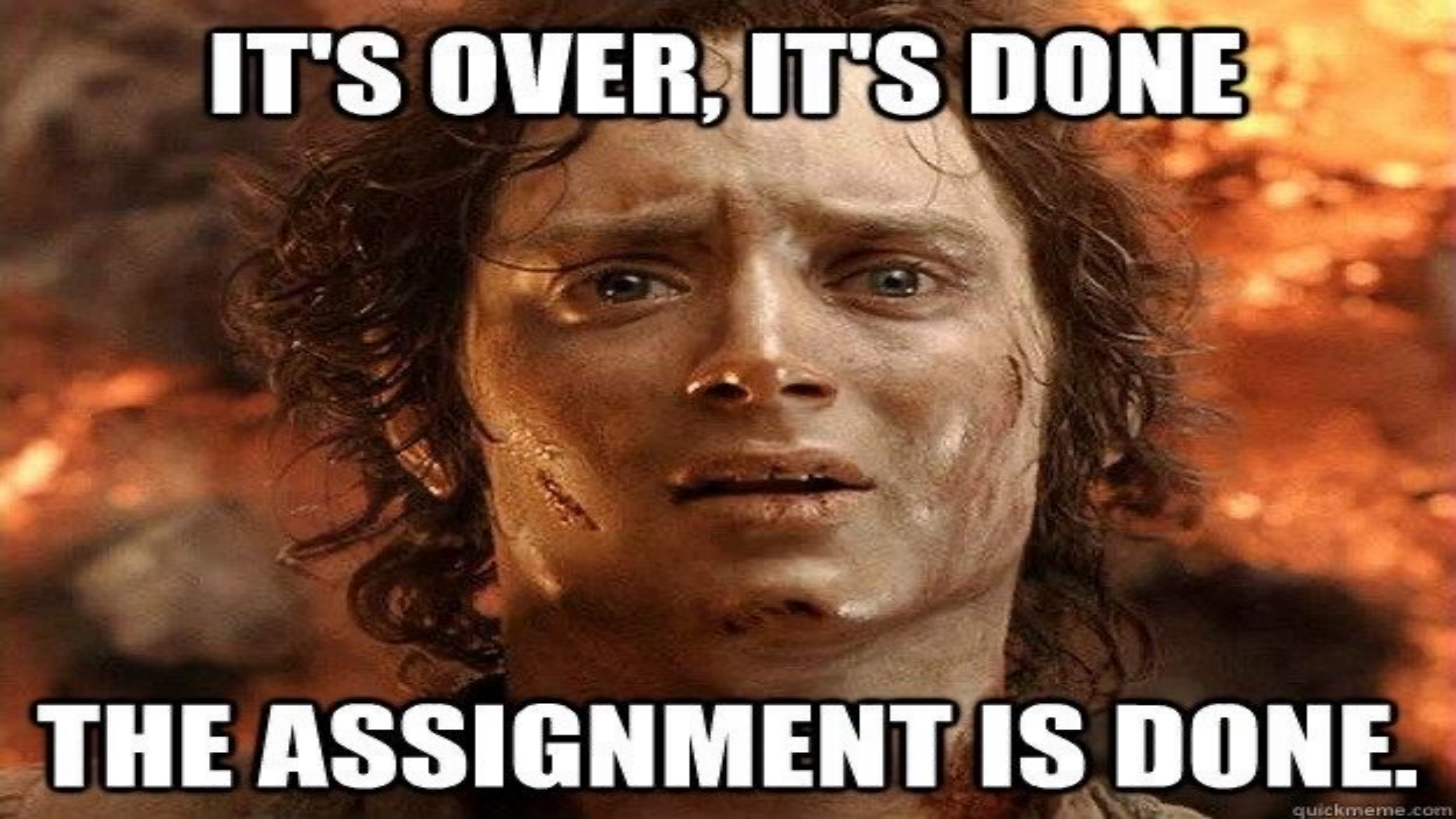
upGrad

Assignment-Endnote

Zip-file

What to keep in mind

- Add comments after every cell of code. So that we can understand your approach and method.
- Describe the results.
- For subjective answers, use DOC and type on it, if you wish to add images you can. But convert it to PDF before submitting.
- Create only one Jupyter notebook.
- Submit one zip file with the code and the PDF.
- Use StackOverflow for dealing with syntax errors. Rather than being stuck at one place or waiting for someone to resolve your doubts, take action and use the resources available on the internet to save time.
- Post on the discussion forums for resolving any doubts you have
- Finally, write code manually instead of copy-pasting from the in-content notebooks provided. Builds a habit of writing code. It's okay to look and write, but don't just copy-paste under any circumstance. Because of just copy-pasting, a lot of our students have faced difficulties in the past when they had to write some code on their interview.


IT'S OVER, IT'S DONE
THE ASSIGNMENT IS DONE.
quickmeme.com