Note – only pdf format is allowed (save doc as pdf)

Add your name/names of team members to the pdf file name and or zip name (1-2 team members)

Phase 1 – due date End of March 2022

Phase 1.1 - Business and Data Understanding

Each answer – 1-3 lines (preferred in English)

1. What are the business needs you are trying to address

2. What are the objectives of the Data Science project – use supervised learning (regression or classification) and define how your target, y, addresses the business need/problem

3. What kind of data is available

4. How large is your data

5. Main features

6. Can you find a new data set online that you could merge and increase your insights

Phase 1.2 - Exploratory Data Analysis – 2-6 interesting visualizations. Self-explanatory

Deliverables (phase 1.1+1.2): Assignment should be delivered in pdf format to enabley, file name should contain student's name

==========

Phase 2 – due date – Middle of July 2022

Data Preparation

1. Missing values or outliers

2. Dummy variables for categorical variables

3. Other techniques we'll learn – imbalance, sensitivity analysis, etc…

Models

Align to the business objectives, try as many models as possible

Evaluation

Depends on the model types and business objective

SUMMARY

Major findings – what you achieved, barriers, next steps

Deliverables: zip file upload to enabley (zip file name should contains student's name) containing (1) jupyter notebook, (2) html notebook (convert jupyter to html) and (3) summary pdf (containing the phase 1 of course)

FINAL PRESENTATION – VOLUNTARY – LAST LESSON MEETING

- Convert the final pdf doc (phase1 and phase 2) to 10 minutes power point presentation (4-7 slides)

Sources:

https://he.flightaware.com/live/cancelled/today

https://casa.sapo.pt/Venda/Apartamentos/?sa=11&lp=10000&or=10

https://www.audiogon.com/listings?order_by=created_at+DESC&page=1&show_media=false

https://www.cbs.gov.il/he/settlements/Pages/default.aspx?mode=Yeshuv

https://data.gov.il/dataset

https://github.com/awesomedata/awesome-public-datasets

https://github.com/fivethirtyeight/data

kaggle

uci data sets

https://data.world/datasets/classification
https://data.world/datasets/regression