

## Assignment 3 (40%)

### Due: 6 Dec 2019 at 12:00pm

#### *Description:*

This is group project. The project is about applying predictive models to solve business problems. Groups need to study a dataset (choose ONE from the given two datasets), identify business objective(s), visualize data and implement predictive models and respond to the identified business problem(s). Groups need to design, build, test, and fine tune the proposed models with Python in order to produce reasonably good or even excellent results.

#### *Learning goals at business level*

- Identify practical and solvable business objective(s) from dataset(s)
- Analyze data with identified business objectives in mind
- Design and implement predictive models to respond to the identified business objectives
- Explain and present the findings from business perspectives

#### *Learning goals at programming level*

- Python programming
- Packages: Pandas, Matplotlib, Seaborn, Scikit-learn, etc.

#### *Datasets*

Students need to choose ONE dataset from the below two datasets for this assignment. Students may make meaningful and practical assumptions when they study and analyze the dataset. Assumptions need to be clearly stated in the submitted documents. *Please note that you need to create a kaggle account in order to download the dataset from kaggle.*

**Dataset 1:** 17K Mobile Strategy Games

<https://www.kaggle.com/tristan581/17k-apple-app-store-strategy-games/kernels>

**Dataset 2:** Airbnb listings and metrics in NYC, NY, USA (2019)

<https://www.kaggle.com/dgomonov/new-york-city-airbnb-open-data>

#### *Assessment components:*

1. Jupyter Notebook (10%)
2. Presentation (20%)
3. Q&A session (10%)
4. Peer Evaluation

**Jupyter Notebook (.ipynb) (10%)**

Tasks:

- a. Identify practice and solvable business objective(s) from a selected dataset
  - b. Analyze the dataset statistically and present the findings and results
  - c. Data cleansing: clean up irrelevant data (could be either rows or columns or both) in the dataset
  - d. Create a predictive model as a baseline and 2 better models (so at least three predictive models)
  - e. Evaluate the models, and choose the best model (explain why)
1. Write down the selected dataset, group number, group member's names and student IDs in the Text cells at the top of the notebook page.
  2. Show clearly the workflow in the submitted Jupyter Notebook document (.ipynb)
  3. Suitable text annotation for readability  
*Do not need to write too many paragraphs / words in the notebook file as you are asked to present in the video.*

**Presentation (20%)**

1. Create PowerPoints slides or other material that helps to present
2. Make a video presentation (in mp4 format) to show the workflow (including the identified business objectives), interesting findings (how the findings respond to the business objectives), and specific techniques (e.g., other packages, classes, functions etc.) to highlight.

**Q&A Session (10%)**

1. This session will be held on **9 Dec (Mon) at 2pm**. (Inform us as early as possible if there is time clash.)
2. A few questions will be released through Canvas.
3. The questions are related to your presentation and your program. Therefore, the questions are different for different groups.
4. Submit the answers within 1.5 hours. (9 Dec 3:30pm)

**Peer Evaluation**

1. A peer evaluation will be released after the Q&A session.
2. The deadline is **10 Dec (Tue) 11:59pm**.
3. Your individual score will be calculated based on this formula:

$$\frac{\text{your evaluation score}}{\text{highest score in your group}} \times \text{Project Score}$$