HONG KONG DESIGN INSTITUTE 香港知專設計學院

**Project Brief**

| | Department of Digital Media | | |
|---|---|---|---|
| Course / Level | 114102 - 2 | | |
| Module | Digital Data | Lecturer | M. C. Sun |

## Project: Collect Open / Real-Time Data, Visualization and Analysis

**Objectives**
This project is designed to enable students

1. produce application with middleware (driver) and NoSQL database.
2. visualize data from open databases and
3. obtain insight / actionable items

**Brief**
You are required to collect open data and real-time data.

Part 1: Jupyter notebook Data Analysis and suggestion of actionable items

- Download the **Top 200 common passwords by country 2021 database** from www.kaggle.com
- Manipulate and rearrange the data if necessary
- Visualize the data using 8 or more charts using Python programming in Jupyter notebook.
- The sunburst chart, heat map, and pair-plot must be used.
- 1 or more 3D chart is essential.
- 1 or more map, such as choropleth map in plotly should be displayed.
- Analyze the charts (and data) which may reveal some facts to us.
- Provide insights and suggest actionable items.
- (You may add other related data set(s) to enrich your insights and suggestions.)

Part 2: Real-time data processing and visualization, in Jupyter notebook.

- Collect and store real-time data using the API of **HK Accident and Emergency waiting time** (of Hospitals) in NoSQL database (e.g., MongoDB).
- The data collection duration should be 3 or more days, within November and/or December.
- The collection frequency should be every 15 minutes or less.
- Create Jupyter Notebook to read data into a Pandas dataframe.
- (You may export the data, using Mongo Compass, to a json file first.)
- Process and visualize the data.
- Produce 3 or more charts.
- You are encouraged to use python 3D visualization techniques too.
- Analyze the charts (and data) to reveal some facts.
- Provide insights / comments / suggestions.

Items should include:
- Exported collection(s) of the open data / samples of real-time data, from MongoDB
- Jupyter Notebooks that visualize and analyze the data sets, with summary, conclusions and so on in Markdown.
- Demonstrate data collection process and present all results / insights, in a video.
- Upload everything to Moodle 1 week after the last lesson.

**Evaluation Criteria**
1. Techniques
   a. Database (MongoDB) (10%)
   b. Visualization (Jupyter) (55%)
   c. Statistics and Data Analysis (Jupyter) (30%)
   e. Demonstration Video Contents, Skills, Audience Experience (5%)

**Deadline of submission: 1 week after Workshop 13 (Dec 8)**

Remarks:

5% of final mark will be deducted for the submission not fulfilling the requirement of submission format.

Late submission penalties

- Assessment submitted within 1 week after the deadline, a penalty of 10 marks would be applied. (e.g. a mark of 67 would become 57 after the penalty)
- Submission after 1 week (on the 8th day) will not be accepted. Therefore, zero grade would be given.
- For the failed case, resubmission within 1 week would be allowed and a maximum of 40 percentage marks would be given for the concerned assessment.