DS102
Project Brief

(May 2019)

The Brief

- Project Overview & Task
- Grading Criteria
- Timeline
- Project Policies

Project Overview

The **DS102 Project** offers you an opportunity to apply your **Exploratory Data Analysis (EDA)** skills to gain insights on **real-world** data.

HACKWAGON
•ACADEMY•

Project Overview

STEP 1

Pick a dataset or a group of (not more than 3) datasets from one of the following sources (more in Appendix A):

- Government Portals
- Open-sourced datasets
- Web-crawled datasets

HACKWAGON
•ACADEMY•

Project Overview

STEP 2

Download and clean the data that you have selected. Perform **Exploratory Data Analysis** and document any **insights** you have uncovered.

HACKWAGON
• ACADEMY •

Project Overview

STEP 3

Submit a **report** in the form of **one (1) Jupyter Notebook** on the insights you have uncovered.

HACKWAGON
• ACADEMY •

Project Overview

You can submit the project as a team of **2 to 4 members**. Each team member should have a **significant contribution** to the project. Larger teams are expected to submit larger projects.

HACKWAGON
• ACADEMY •

The Brief

- Project Overview & Task
- Grading Criteria
- Timeline
- Project Policies

Grading Criteria

CAUTION!

The following are just *guiding questions* for each section of the report. You **do not** need to answer every question. A good report should be clear and coherent.

HACKWAGON
• ACADEMY •

Grading Criteria

I. **Executive Summary**

(3 marks)

II. **Problem Statement & Dataset Selected**

(3 marks)

III. **Methodology, Insights & Evaluation**

(14 marks)

HACKWAGON
• ACADEMY •

Executive Summary [3m]

Guiding Instructions & Questions

Pick out the key statements from your report and condense it into a 1 - 2 paragraph summary. What was the introduction, hypotheses, methods and results from your research? What discussion points have surfaced as a result of your research?

Problem Statement & Dataset Selected [3m]

Guiding Questions I

- What research topic did you select?
- Why is this topic interesting to you?

Guiding Questions II

- What dataset did you pick for analysis?
- What was the structure of the dataset (CSV file, API calls, web-scraped etc.)?

Methodology, Insights & Evaluation [14m]

Guiding Questions I

- What steps did you take to clean the dataset / fill missing data points?
- What EDA techniques did you use? These can include using descriptive analysis and visualisation tools used.

Methodology, Insights & Evaluation [14m]

Guiding Questions II

- What insights have you uncovered from the analysis? Did they support your initial hypothesis?
- What were discussion points surfaced after completing this project?
- What other possible future research efforts could be performed by extending this project?

The Brief

- Project Overview & Task
- Grading Criteria
- Timeline
- Project Policies

Project Timeline

Week 2 (Today)

Project Released. Decide your project topic & evaluate datasets. Also form teams if you'd like to do this in a team.

Before Week 5 class

Submit Proposal. Proposal should cover the **Problem Statement & Dataset Selected** segment. Submit the proposal in the form of **one Jupyter notebook**. The proposal is **optional.**

HACKWAGON
•ACADEMY•

Project Timeline

Before Week 7 Class

Submit your final project as **one Jupyter Notebook to eLearn.** Please add in all names of the team members in your team. Retain your datasets as we might request them from you.

Any late submission after Week 7 class starts will be penalised, with **50% of the marks deducted.**

HACKWAGON
•ACADEMY•

The Brief

- Project Overview & Task
- Grading Criteria
- Timeline
- Project Policies

Project Policies

Grading

The results of your project is worth 40% of your final grade. Marks are awarded based on how well the report has addressed each of the 3 key segments.

HACKWAGON
• ACADEMY •

Project Policies

HWA Honour Code

- You may use any existing code or libraries or consult any online and offline references for your project. However, they **must be cited**.
- If you are using an alternative data source, it **MUST be public**. The teaching team has the right to reject any projects that use proprietary artefacts (datasets / techniques).
- You are **NOT ALLOWED** to look at another DS102 project and use their code in your project.

HACKWAGON
• ACADEMY •

Appendices

Appendix A

Government Portals

https://www.data.gov/

https://data.gov.sg/

https://www.gapminder.org/

https://data.worldbank.org/

Dataset Search

https://toolbox.google.com/datasetsearch

Open-Source Datasets

https://www.kaggle.com/datasets

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

https://www.kdnuggets.com/datasets/index.html

HACKWAGON
• ACADEMY •