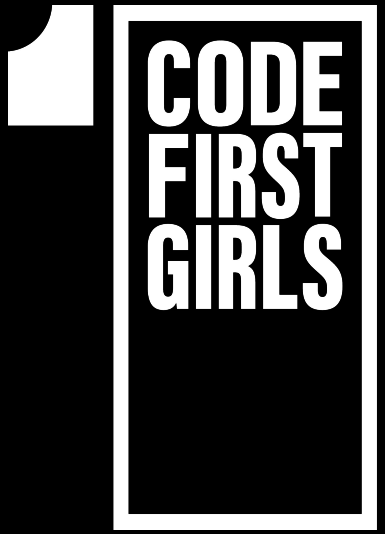
PROJECT WORKSHOP

LESSON 4



CFGDEGREE → DATA SCIENCE MODULE

AGENDA

01 Identifying a suitable project

02 Identifying your strength

03 Project examples

04 Project criteria

05 Form groups

WHAT vs HOW

- Choosing the RIGHT project is already 40% of your success.
- Don’t worry about ‘HOW’, spend time considering the right ‘WHAT’!
- Love your project idea, the ‘HOW’ will come or can be researched and learnt.
- This project would form the first significant entry in your coding projects portfolio and will be used in interviews.

PROJECT

1: CHOOSE SOMETHING YOU ENJOY WORKING WITH

1. The first thing you should think about is what you enjoy doing in life in general.

1. The easiest place to start looking for inspiration is your hobbies. What do you do in your free time?

PROJECT

2: START SMALL

1. Start building a project that is really simple but still allows you to learn new things. Do it well.

1. The last thing you want is to set yourself too ambitious goals and lose your motivation once things get tricky.

PROJECT

3: AIM SLIGHTLY PAST YOUR CURRENT SKILL LEVEL

1. The point of picking the perfect coding projects is to choose something that reaches just past your current skill level.
2. It is the only way to push your limits and help you learn new things.
3. Because the last thing you want is to get stuck in a loop of projects where you repeat the same things over again.

PROJECT

4: PICK A PROJECT TO SOLVE A PROBLEM

1. The goal of any analytics project is to create something useful and helpful.

1. Your analysis should discover or showcase some useful findings to provide insightful data into specific questions.

QUICK EXPERIMENT

USING CHAT WINDOW POST ID FOR YOUR PREFERRED AREA

1. Aerospace and Defence
2. Asset and Wealth Management
3. Automotive
4. Banking and Capital Markets
5. Business Services
6. Capital Projects and Infrastructure
7. Charities
8. Chemicals
9. Education
10. Engineering and Construction
11. Film and cinema
12. Financial Services
13. Forest, Paper and Packaging
14. Government and Public Services
15. Healthcare
16. Hospitality and Leisure
17. Insurance
18. Manufacturing
19. Media and Entertainment
20. Mining and Metals
21. Oil and Gas
22. Pharmaceutical and Life Sciences
23. Power and Utilities
24. Real Estate
25. Retail and Consumer
26. Technology
27. Telecommunications
28. Transport and Logistics

TRANSFERABLE SKILLS & KNOWLEDGE

IMPORTANT

1. Some of you may have backgrounds in different industries, so you already may have a bunch of very useful skills that would be a great base knowledge for your project

1. It is also possible that you may have some very well developed soft or non-technical skills that are HUGELY important in the tech world. In fact, some people say that you need 50% tech skills and 50% soft & non-tech skills to be successful in IT.

1. By transferrable non-technical skills we mean such experience as HR, Finance, Journalism, Sales and so on. When we say 'soft skills', we mean leadership, communication, presentation skills and many more.

SWOT Analysis

DEFINITION

Strengths: What is your team really good at? What are the tech components or libraries that you know the best?

Weaknesses: What are some things that your team is not very good at, where does the team lack knowledge?

Opportunities: Can we leverage non-tech skills and knowledge of our team members to build a meaningful project?

Threats: What are some external factors — how much time can we devote to the project, how much each member contributes? Can we meet the deadline and all criteria?

- Writing a good SWOT analysis starts with brainstorming, or brainwriting.
- Get your team together and start coming with SWOT points

SWOT Analysis

EXAMPLE

The Data Analysis Process – high level definition

Step 1: Frame the problem	Formulate exact questions that we are trying to answer or problems that we are trying to solve
Step 2: Collect the raw data	Data gathering: series of data captured for a specific time period, relevant records, values etc.
Step 3: Prepare the data for analysis	Build a DB with data arranged in tables. 'Clean' the data, enrich it with required additional values etc.
Step 4: Explore the data	Review, understand the data and perform **summary statistics with descriptive analysis**.
Step 5: In-depth analysis	Usually **Machine Learning, AI analysis using regression, mathematical model building -- predictive analysis**.
Step 6: Communicate results	**Data visualisation, findings summary, interpretation of results , recommendations, conclusions**

Step 1:
Frame the problem

Formulate exact questions that we are trying to answer or problems that we are trying to solve

Step 2:
Collect the raw data

Data gathering: series of data captured for a specific time period, relevant records, values etc.

These steps are going to be the hardest!

- Spend time refining WHAT EXACTLY you want to analyse
  (limit it to core 3-4 questions)

- **Do a research** – where are you going to get data for your analysis?
  It is probably going to be some historical data:
  - are we using API,
  - are we downloading CSV or XLSX from somewhere?

EXAMPLES

ANALYSIS PROJECTS

- Do a stock analysis to explain why some stocks went through a significant drop in price.
- Study the factors contributing to air pollution in a given city.
- Study the likelihood of a worsening or improving diabetes in children
- Identify the common skills and qualifications of the top-performing employees in a company.
- Do data mining for increasing and predicting sales in tourism.
- Do data mining for increasing and predicting sales in tourism.
- Credit card fraud detection. (Detect anomalies)
- Do a segmentation of a customer data set from an e-commerce website.
- Analyze a TV show popularity.

WHATEVER YOU DO YOU NEED TO FIND **GOOD DATA**

PROJECT CRITERIA

REQUIREMENTS

1. **Project to be delivered in groups 3-5 people**
2. **Project has a clear objective:**
    a. solves a clearly defined problem
    b. Provides useful insights into a data set that are not obvious to a naked eye
3. **Projects is built and completed within given time frame:**
    a. NB: you cannot ‘reuse’ any past projects, it has to be a new one
    b. A crucial part of your project would be to find suitable, relevant datasets (online sources, data libraries, APIs etc.)
    c. Note that your data availability may dictate the topic you would be choosing.
    d. There are many good data resources online. Do a thorough research to find suitable data.

PROJECT CRITERIA

MUST HAVE

1. **Python code script in Jupyter Notebook:**
    a. Has a clear structure, with clearly defined sections: **loading data, cleaning data, transforming data, analysis, visualization, reporting.**
    b. Separate **data files submitted** : csv, xlsx, txt etc. Note that if you loaded data from some API, you still need to provide a file where your data is saved.
    c. An instructor or assessor should be able to run submitted scripts with the files provided and get expected results.
2. **Use at least one API to fetch data**
3. **Use of key scientific packages: Pandas, NumPy, MatplotLib**

PROJECT CRITERIA

GOOD TO HAVE

1. **Machine Learning**
2. **SciKit Library**
3. **SQL database**

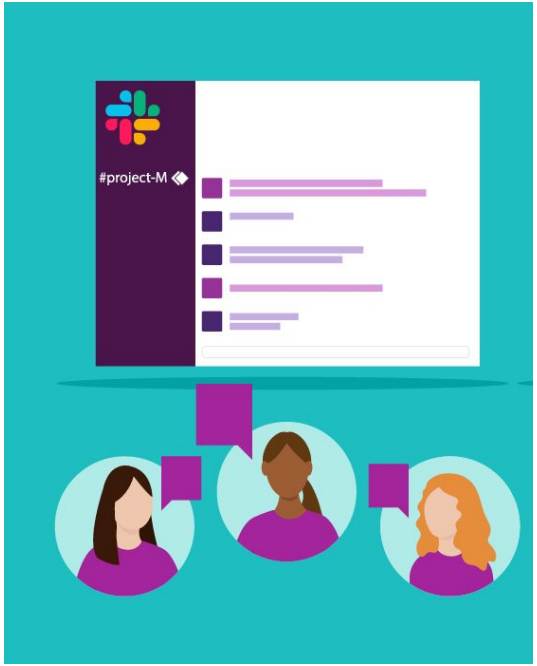
PROJECT CRITERIA

SUBMISSION

1. **PDF document** (as per guidelines) with clear project specification
2. **Source code** for the project to be shared **via GitHub** with your instructor:
   a. Source code (Jupyter notebook file)
   b. Data files
   c. README file with clear instructions how to execute the code
3. **PPT slide deck** (2-3 min length max) with key points for presentations.

SPLIT INTO TEAMS

1. Aerospace and Defence
2. Banking, Trading, Finance
3. Business Services
4. Charities
5. Education
6. Environment, recycling
7. Film and cinema
8. Government and Social Services
9. Healthcare
10. Hospitality and Leisure
11. Insurance
12. Media and Entertainment
13. Pharmaceutical and Life Sciences
14. Retail and Consumer
15. Telecommunications
16. Transport and Logistics



THANK YOU!