SUMMER ANALYST CHALLENGE

TECH INTERN HACKATHON

neueda TRAINING

TOPICS

- Hackathon Goals – Data Analysis
- Hackathon Goals – Technical
- Key Deliverables
- Example Data Sets
- Hints and Tips FOR A successful Project

SUMMER ANALYST CHALLENGE

- End the final day of the 4-day program with a short informal presentation to instructors and peers on the work that has been done over the course of those 4 days training.
  - Build a team bond for the Summer Analyst Challenge to come.
  - End the first phase of training with a clear deliverable and achievement.

- Summer Analyst Challenge tasks groups with working as a team to:
  1. Use Python to clean, wrangle and combine multiple financial datasets.
  2. Prepare a highly visualized story using python visualization tools such as Plotly, jupyter notebooks that processed data to present insights and draw conclusions.
- Analysts or Teams moving into roles with more coding can reach further and put more emphasis on data processing and analytics with Python.
- Analysts or Teams moving into roles with less coding can reach further and put more emphasis on storytelling with data and visual reporting.

HACKATHON GOALS – DATA ANALYSIS

- Select one or more public domain datasets
- As a group decide to take one or more of the following approaches:
  - Decide on some "questions" you will ask of the data. The goal of your analysis will be to answer these questions.
  - Decide on some "hypotheses" you have about the data. The goal of your analysis will be to prove or disprove these hypotheses.
  - Decide on a "story" that you will tell with the data. The goal of your analysis will be to highlight the key points in your story.

HACKATHON GOALS – TECHNICAL

- Use cloud Jupyter Lab environments for storage and file sharing ***whenever possible***
  - Try to avoid sending files to each other and losing track of versions!
  - Feel free to use revision control such as git if you have experience with it – but don't let this slow you down!
- Demonstrate your abilities with the Python / Pandas / Jupyter environment for cleaning, sorting, shaping, merging etc.

HACKATHON GOALS – TECHNICAL (CONTINUED)

- Demonstrate your skills with visualisation
- Take advantage if you see an opportunity to bring in any skills or experience that you have e.g.
  - A Python library you've used in the past
  - Any statistical or analysis techniques you've used
  - Any opportunities to bring in machine learning techniques

KEY DELIVERABLES

1

A Python Jupyter notebook that cleans, shapes, joins, visualises your data as you see fit.

- Make use of markdown cells to really "present" your notebook.

2

An engaging presentation – consider the techniques covered during presentation skills module.

- Use a small slide deck, your python code & a data dashboard to present.

EXAMPLE DATA SETS

- Kaggle is the go-to place for interesting data sets.
  - Here are some examples to help you get started

- Electic Vehicles
- US Consumer Finance Complaints
- Electric Car Sales
- Insurance Data
- Supermarket Sales

CHALLENGE MILESTONES

Milestone One – End of Day 1

- **Have chosen a domain for your final project:**
  - For example, mortgage defaults over last 3 years, electric vehicle sales, plant-based diets by age groups, etc.
- **Have 3 or 4 sentences that describe the following:**
  - Your hypotheses i.e. what you are trying to prove / disprove
  - Your data sets – types of data, sources of data
- **Have a simple notebook ready** that imports your data into a pandas DataFrames(s)

Milestone Two – End of Day 2

- **Have cleaned and prepared your data.**
  - Replace NaN, generated new data, uses Imputers, etc.
- **Joined your datasets**
  - merged, joined, concatenated your multiple datasets into a single Data Set
- **Export to csv**
  - Your new dataset single csv or excel spreadsheet, ready for tableau

Milestone Three – End of Day 3

- **Have created the narrative of your story**
  - Background, context
  - Sources of data
  - Hypotheses
  - How are you going to support your hypotheses
- **Have decided on the types of charts, color schemes, univariate, multi variate data you are going to present**

Milestone Four – End of Day 4

- **Planned your presentation**
  - Notebook
  - PPT presentation
- **Decided who is responsible for each part of your final project**
- **Created a plan to ensure you meet your deadline**
  - Daily check-ins, create a group zoom room or google chat etc.
- **Solicited some final tips, advice from your instructors.**

HINTS AND TIPS FOR A SUCCESSFUL PROJECT

- Use your Python skills to merge multiple data sets together into a single file
  - This should include cleaning data, removing duplicates and finally exporting to a spreadsheet or csv file
  - This final file is what you should be using as the data source to your visualizations
- Try to include different types of data
  - **Time series data** – measurements of the same data taken over different time periods
  - **Geographical Data** – data that has countries, states, cities, zip codes etc. Python has some very geospatial packages that produce very interactive and impressive maps
  - **Derive some data** – Again, using your python and pandas skills, use some features like group by, pivot, melt and also some simple statistics such as mean, std, or even your own had written ones, to derive some new values such as moving averages, correlations etc
- This is not a big data exercise!
  - Whilst it might be tempting to munge and wrangle millions of rows of data, the objective of this exercise is to
    - Use python to prepare some data to be ingested into a tableau dashboard
    - Tell a story in an engaging and captivating manner
    - That is finance related