Capstone: Group

This project is an opportunity for you to consolidate what you have learned during the course and apply it to a business problem that you work on with others.

You will have **three days** of class period to work on this project. Your group will turn in one version of all files that you use.

Scenario

You work for **WOW!**, a clean energy private equity firm based in Chicago. So far their investments have been in U.S. hydropower and they are now thinking about investing in U.S. wind power.

Your manager comes to speak with you and tells you that a team of data analysts is being formed **to assess the U.S. wind turbine market in a short turnaround.**

The team needs to present its findings to the investment committee, in time for their quarterly investment decisions.

Learning Objectives

- Conduct market research on opportunities for **WOW!** in the U.S. wind turbine market: **Should WOW! invest in the wind market? If so, who should they approach to fund?**
- Utilize the data analytics workflow within an agile development framework.
- Transform data into useful information to help support business decisions.
- Gain experience with project management and associated tools.

Deliverables

Submit a 20-minute client facing presentation containing:

- An overview of your analysis for the Investment Committee.
- Retrospective presentation on Project Experience including:
    - What you have learned from this project.
    - Any additional files that you used as part of your analysis.
    - [Optional] Screenshot (or share link) to your team's Trello Board.

Data

You can get the data here:

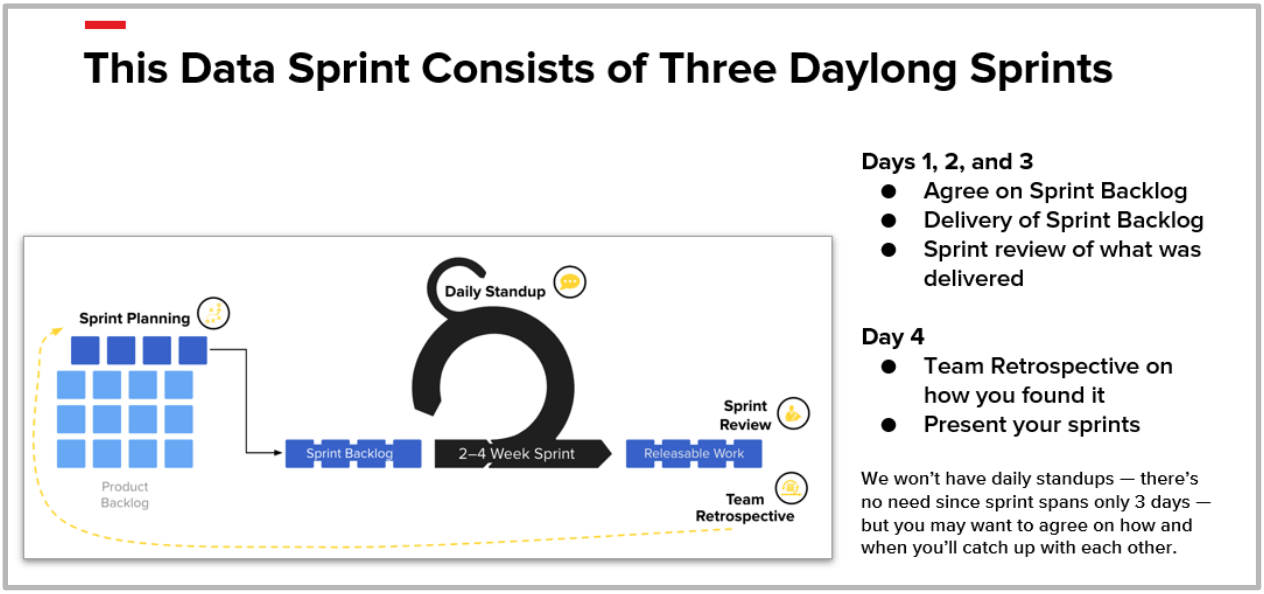
- **Main file:** Data on wind turbines in the U.S. (location, which plant they are in, how much power they generate, etc.) (wind_turbine_20220114.csv). It is from the U.S. Wind Turbine Database, and includes information on 70,808 turbines covering 44 states (plus Guam and Puerto Rico). The data are used by government agencies, scientists, private companies, and citizens for a variety of analyses.
- Data dictionary for the wind turbine data (wind_turbine_20220114.csv).
- Data on operators within the U.S. energy space. Data is from here (2022: EIA-923 January 2022 zip file). Hints:
    - Look at 'Page 1 Generation and Fuel Data' sheet to see Operators
    - The file contains a data dictionary on the last sheet 'Page 7 File Layout'
- The above data is also in PGAdmin for you to analyze, specifically:
- wind_turbine_20220114 table: contains the wind turbines csv file above
- eia923_operators table: contains the data in the 'Page 1 Generation and Fuel Data' sheet in the operators data.
- Reminder: here are the login details for PGAdmin
    - https://analyticsga-euwest1.generalassemb.ly/browser/
    - **Username**: analytics_student@generalassemb.ly
    - **Password**: analyticsga (Check ☑ Save password)
- How many wind turbines are contained in the U.S. Wind Turbine Database?
- Average wind speeds in the U.S.A

Skills You'll Use

This lab will let you try out the key skills we have learned throughout the course across the data analytics workflow, including: Framing questions with stakeholders, cleaning and processing data, visualizing data, communicating your data, and modeling your data. And all this by working with others in an agile team.

It will require you to utilize a combination of Excel, SQL, Python, Power BI and/or Tableau skills as you work through the data analytics workflow to analyze a possible investment in alternative energy. However, in this sprint, you will also work within an agile framework.

Structure of the Sprint



- The data sprint will be a three-day 'sprint simulation': A hypothetical scenario which is based on a real-world alternative energy industry.
- You will be put into groups and will work in those breakout groups throughout the sprint.
- You will run a sprint across the first three days, and then write up the sprint in the afternoon of day three and present your findings and what you have learned from the process in the morning of day four.
- Instructor Team will be playing the role of product owner and scrum master, and will help prioritize your work and coach you. The product owner has already created a product backlog and estimated the effort it will take for each item.
- You need to agree on the sprint backlog (high-level milestones) to complete with your group each day; you will also need to agree on a strategy for the day and a plan to divide up the work amongst you.
- You will conduct a sprint review with your product owner at the end of days one and two, to show them what you have managed to complete from the sprint backlog.
- We ask that you approach these three days with a creative analytical mindset!

Background

- Renewable energy is energy that is collected from renewable sources that are naturally replenished on a human timescale. https://en.wikipedia.org/wiki/Renewable_energy
- A wind turbine is a device that converts the wind's energy into electrical energy. Smaller wind turbines are used for applications such as battery charging and to power traffic warning signs. Larger turbines can contribute to a domestic power supply. https://en.wikipedia.org/wiki/Wind_turbine
- You have been tasked with analyzing the wind turbine industry and outlining recommendations for a strategy for investment in the renewable energy industry.

Daily Breakdown of Tasks and Deliverables

Preliminary Work

Project Introduction (90 minutes)

- Go through what the sprint involves, and the scenario.
- Discuss how you will divide responsibilities for delivering as much value from the backlog as you can in two days.
- Clarify any questions you have on how the sprint will work.

Preparation Session (90 minutes)

- Research the renewable energy industry and wind turbines to understand what wind turbines are, how they work (at a high level), and why they are used. You will need this to be able to effectively analyze the data. For your research, read the links in 'Background' above, and also these links:
    - Electricity explained: electricity generation, capacity, and sales in the U.S.
    - Wind Resource and Potential
    - Our World in Data: Energy Consumption by Source (scroll down to see other data and graphs)
    - Key trends in the U.S. wind market: Land-Based Wind Market Report
    - **Stretch:** Use anything else you know of or can find!
- Investigate the data sets provided and review the data dictionaries:
    - See the links above
    - See also **Tips For Inspecting Your Data Sets** below
    - **Stretch:** Look for other data sources (or use ones you already know about); you likely won't use these until the second sprint on day two
    - If you have time, conduct some initial EDA (descriptive statistics, look for missing data, visualizations)
- Review the scrum and agile workflow, and how we will use them in this data sprint (see slides).
- Review the product backlog, clarify anything you need with the product owner (Instructor) and be prepared to agree upon the sprint backlog for day one within the first hour of day one.

Day 1

Agree on the sprint backlog for sprint one (1 hour)

- Work out what you can accomplish as a team today, and any help you may need
- Ask the product owner for any questions to clarify items in the product backlog
- See the **tips for creating your sprint backlog** section below for help with structuring
- Agree on the sprint backlog with the product owner for sprint one (day one)

Team Work Tasks (rest of the day except for the sprint review)

- Complete your sprint backlog utilizing Excel, SQL, Python, Tableau and/or Power BI to engage in the DA workflow.

Sprint Review (30 minutes)

- Complete a 10-minute show-and-tell of the work that was completed during the day with the product owner.
    - Summarize the team's findings and insights and any tasks that still need to be completed. These can be notes, and can be in any format you choose. Review visualizations and dashboards that were created.
    - Discuss any challenges or struggles that you experienced and how you overcame them.
    - Discuss further steps that need to be taken and any part of the sprint backlog that was not completed.
- Spend no more than 20 minutes preparing for the show-and-tell
- Submit your show-and-tell
- You do not need to submit any other files at the end of today

Day 2 / Day 3

Retrospective (30 minutes)

- In your team, reflect on day one / day two and answer five questions:
    a. What went well?
    b. What didn't go so well?
    c. What have you learned?
    d. What still puzzles you?
    e. Do you need to change your ways of working at all for day two?
- Discuss and agree on the structure of day two / day three, taking into consideration your reflections.

Agree on the sprint backlog for sprint two (30 minutes)

- Work out what you can accomplish as a team today, and any help you may need
- Ask the product owner for any questions to clarify items in the product backlog
- Agree on the sprint backlog with the product owner for sprint two (day two)

Team Work Tasks (rest of the day except for the sprint Review)

- Complete your sprint Backlog utilizing Excel, SQL, Python, Tableau and/or PowerBI to engage in the DA workflow.

Sprint Review (30 minutes)

- Complete a 10-minute show-and-tell of the work that was completed during the day with the product owner.
    - Summarize the team's findings and insights and any tasks that still need to be completed. These can be notes, and can be in any format you choose. Review visualizations and dashboards that were created.
    - Discuss any challenges or struggles that you experienced and how you overcame them.
    - Discuss further steps that need to be taken and any part of the sprint backlog that was not completed.
- Spend no more than 20 minutes preparing for the show-and-tell.
- Submit your show-and-tell.
- You do not need to submit any other files at the end of today

Day 4

Retrospective (30 minutes)

- In your team, reflect on day two and answer five questions:
    a. What went well?
    b. What didn't go so well?
    c. What have you learned?
    d. What still puzzles you?
    e. Do you need to change your ways of working at all for day three?
- Discuss and agree on areas of improvement that you can use for working together on day three.

Team Work Tasks

- Prepare a presentation that includes your team's insights and recommendations actions for the WOW! investment committee:
    - Review the team's findings, insights and recommended actions.
    - Discuss any challenges or struggles that you experienced and how you overcame them.
    - Discuss further steps that would be taken in the next sprint.
- Tip: You can pick a slide template from https://slidesgo.com/ and download it.

Presentations

- Each team will deliver a 25-minute presentation:

    a. Proposal for the WOW! investment committee (15 minutes)
    b. Lessons learned as a team from doing the project (5 minutes)
    c. Q & A (5 minutes)
- The intended audience of the presentation is technical and non-technical.
- Each team member will contribute to the delivery of the presentation.
- The presentation pdf will be submitted by Day 4 at 9am. All other files will be submitted by 11.59pm on Day 3. Include a doc that summarizes all submitted materials.

Tips For Creating Your Sprint Backlog

Spend time looking at all the items in the product backlog. Their priority has been set by the product owner, but you can discuss this with them if you disagree with any of them.

The estimated relative priority of each item (Story Points) has been set by the product owner, who talked to various individuals in your department as they defined the backlog.

You need to take the product backlog, clarify anything you need to (which may involve the product owner defining some items in more detail, or even splitting them out into multiple items), and then agree with the product owner which ones you'll do in your sprint.

For each item in the product backlog, consider:

- Is the item well written? Is it clear what's required?
- What would you need to do in order to complete the task?
    - What data would you need?
    - What data wrangling would you need to do?
    - What tools would you use? Excel, Python, SQL, Tableau, Power BI, etc.?
    - What are some possible issues you might face?
- Could you divide and conquer, or would you all need to work on it together?
- How long do you think it would take? Is the relative effort listed in the product backlog right?
- Do you think the item's priority is right?

Tips For Inspecting Your Data Sets

Start with ***loading the data*** into Excel or Python, or examine it using SQL, and get familiar with what you have (and what you don't have).

When inspecting the data sets:

- What does each column mean?
- Which features (columns) do you think will be the most useful?
- Which features (columns) do you think you won't need?
- Is there missing data? If so, how could you handle it?
- What other data would you like to get, to help your analysis? Is it easy to find and get? If so, is it complete?
- Conduct descriptive statistical analyses in Excel or Python.

Work out what data cleansing / wrangling is required

- Which data source(s) will you use to answer each question in your sprint backlog?
- What cleaning do you need to do to the data? For example, renaming columns, joining data, and handling nulls.

Create data table relationship(s) in order to run queries in SQL.
Conduct visualizations and dashboards in Excel, Tableau, Power BI or Python.

Explain your data wrangling process (if any) when you present to the investment committee. Note any important new columns you created in your data tool and any steps you took to clean the data, change data types, handle missing data, merge data, etc.
Include comments throughout your work.

Good luck and have fun!