**Capstone 2: Project Ideas**

**A. Battery Remaining Useful Life**

The Hawaii Natural Energy Institute examined 14 NMC-LCO 18650 batteries with a nominal capacity of 2.8 Ah, cycled over 1000 times at 25 °C with a CC-CV charge rate of C/2 and a discharge rate of 1.5 C. The dataset includes voltage and current patterns observed during each cycle, which can be used to forecast the remaining useful life (RUL) of the batteries. It provides a summary of the performance data for 14 batteries with the following variables:

1. Cycle Index
2. Discharge Time (s)
3. Decrement 3.6-3.4V (s)
4. Max. Voltage Discharge (V)
5. Min. Voltage Charge (V)
6. Time at 4.15V (s)
7. Time Constant Current (s)
8. Charging Time (s)
9. RUL

The dataset is built by IGNACIO VINUALES, and the process is described in <https://github.com/ignavinuales/Battery_RUL_Prediction>. The data is available for download on <https://www.kaggle.com/datasets/ignaciovinuales/battery-remaining-useful-life-rul>. I plan to develop my own Remaining Useful Life predictor from this data. The downside is that the dataset does not have missing values, providing limited experience in data wrangling. However, this project is closest to my previous research work as a PhD student, making it a strong bridge to my new career.

**B. Amazon Products Sales Dataset 2023**

Amazon is a multinational American technology company with diverse business interests, including e-commerce. In its e-commerce operations, Amazon manages all aspects of the process from shipping and pricing to customer service and returns. I will work mainly on the largest .csv file, "Amazon-Products.csv", which contains data on over 500,000 products sold by Amazon. The dataset includes the following features:

1. name: The name of the product
2. main\_category: The main category to which the product belongs
3. sub\_category: The subcategory to which the product belongs
4. image: The product image
5. link: The Amazon website reference link of the product
6. ratings: The ratings given by Amazon customers
7. no of ratings: The number of ratings for the product
8. discount\_price: The discounted price of the product
9. actual\_price: The actual MRP of the product

The dataset, scraped from the Amazon website by ABHISHEK SHARMA, is available on <https://www.kaggle.com/datasets/lokeshparab/amazon-products-dataset>. I plan to predict the best price to sell a product based on its features. One benefit of this dataset is the presence of multiple missing values, which can be addressed. Additionally, it can be restructured to estimate the discount price for products currently without discounts. This project is my second priority, as I am aiming for a career in large corporations. The downside is that the specificity and quantity of products might limit my ability to integrate additional data at this stage of my learning.

**C. Mobile Price Prediction**

This Smartphones dataset, scraped from the web by INFORMROHIT1, consists of 210 rows and 26 columns. The data is available on <https://www.kaggle.com/datasets/informrohit1/smartphones-dataset> and includes the following columns:

1. brand\_name
2. model
3. price
4. avg\_rating
5. 5G\_or\_not
6. processor\_brand
7. num\_cores
8. processor\_speed
9. battery\_capacity
10. fast\_charging\_available
11. fast\_charging
12. ram\_capacity
13. internal\_memory
14. screen\_size
15. refresh\_rate
16. num\_rear\_cameras
17. os
18. primary\_camera\_rear
19. primary\_camera\_front
20. extended\_memory\_available
21. resolution\_height
22. resolution\_width

The dataset contains data on different phone models and brands, with several missing values. This makes it suitable for various analyses and prediction models. I plan to build a model to predict the price of smartphones. The dataset's small size and familiarity of the products will allow me to supplement it with additional data from the internet.

**D. USA Real Estate Dataset**

To broaden my learning experience, I chose this common type of dataset to predict house prices. The data was collected from <https://www.realtor.com/> by AHMED SHAHRIAR SAKIB and is available on <https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset>. The dataset includes the following features:

1. brokered\_by
2. status
3. price
4. bed
5. bath
6. acre\_lot
7. street
8. city
9. state
10. zip\_code
11. house\_size
12. prev\_sold\_date

I chose this dataset because it is generic and will help develop my data wrangling skills. Additionally, it might assist in determining the offer price for my first home 😊.

All the data are uploaded in my GitHub and can be located at,

<https://github.com/pparaju/Capstone_Two_Project_Ideas>.