**Three datasets of interest**

**Topic 1**

Title: Prediction of Stock APPL (**Reserved for capstone 3**)

Description: The stock market is intriguing. I’d like to learn and practice skills to understand the stock APPL and predict its performance. Specifically, its historic data and other stocks like AMZN, GOOGL, FB, and FB will be used as features in the model.

Data source: The data of APPL and other technical companies are available from 2005 to 2020 via the link.

<https://www.kaggle.com/datasets/nikhilkohli/us-stock-market-data-60-extracted-features?resource=download>

**Topic 2**

Title: Supermarket store branches sales analysis

Description: I’d like to analyze stores sales and its dependence on stores area and daily customercount.

Data source: In the dataset, You'll get data of different stores of a supermarket company as per their store IDs which for ease has been converted to positive integers.

<https://www.kaggle.com/datasets/surajjha101/stores-area-and-sales-data>

**Topic 3**

Title: Student Performance Analysis

Description: Student’s performance is the key to the education. I’d like to expore the influencing factors in student’s performance and derive a model to predict student’s performance given key features.

Data source: Student Performance Data was obtained in a survey of students' math course in secondary school. It consists of 33 Column Dataset Contains Features like gender, age, size of family, Father education, Mother education, Occupation of Father and Mother, Family Relation, Health, and Grades, etc.

<https://www.kaggle.com/datasets/devansodariya/student-performance-data>

school: Index(['GP', 'MS'], dtype='object')

sex: Index(['F', 'M'], dtype='object')

age: Int64Index([16, 17, 18, 15, 19, 20, 22, 21], dtype='int64')

address: Index(['U', 'R'], dtype='object')

famsize: Index(['GT3', 'LE3'], dtype='object')

Pstatus: Index(['T', 'A'], dtype='object')

Medu: Int64Index([4, 2, 3, 1, 0], dtype='int64')

Fedu: Int64Index([2, 3, 4, 1, 0], dtype='int64')

Mjob: Index(['other', 'services', 'at\_home', 'teacher', 'health'], dtype='object')

Fjob: Index(['other', 'services', 'teacher', 'at\_home', 'health'], dtype='object')

reason: Index(['course', 'home', 'reputation', 'other'], dtype='object')

guardian: Index(['mother', 'father', 'other'], dtype='object')

traveltime: Int64Index([1, 2, 3, 4], dtype='int64')

studytime: Int64Index([2, 1, 3, 4], dtype='int64')

failures: Int64Index([0, 1, 2, 3], dtype='int64')

schoolsup: Index(['no', 'yes'], dtype='object')

famsup: Index(['yes', 'no'], dtype='object')

paid: Index(['no', 'yes'], dtype='object')

activities: Index(['yes', 'no'], dtype='object')

nursery: Index(['yes', 'no'], dtype='object')

higher: Index(['yes', 'no'], dtype='object')

internet: Index(['yes', 'no'], dtype='object')

romantic: Index(['no', 'yes'], dtype='object')

famrel: Int64Index([4, 5, 3, 2, 1], dtype='int64')

freetime: Int64Index([3, 4, 2, 5, 1], dtype='int64')

goout: Int64Index([3, 2, 4, 5, 1], dtype='int64')

Dalc: Int64Index([1, 2, 3, 5, 4], dtype='int64')

Walc: Int64Index([1, 2, 3, 4, 5], dtype='int64')

health: Int64Index([5, 3, 4, 1, 2], dtype='int64')

absences: Int64Index([ 0, 2, 4, 6, 8, 10, 14, 12, 3, 16, 7, 5, 18, 20, 11, 9, 13,

15, 22, 1, 38, 30, 40, 23, 19, 28, 75, 21, 24, 56, 26, 54, 25,

17],

dtype='int64')

G1: Int64Index([10, 8, 11, 7, 12, 13, 9, 14, 15, 6, 16, 18, 17, 5, 19, 4, 3], dtype='int64')

G2: Int64Index([9, 10, 12, 13, 11, 15, 8, 14, 7, 5, 6, 16, 0, 18, 17, 19, 4], dtype='int64')

G3: Int64Index([10, 11, 0, 15, 8, 13, 12, 9, 14, 16, 6, 18, 7, 5, 17, 19, 20, 4], dtype='int64')