**Preliminaries for Data Analysis**

**Instructions:**

Please share your answers filled in-line in the word document. Submit code separately wherever applicable.

Please ensure you update all the details:

**Name: \_\_\_\_Jaidev Chhabria\_\_\_\_\_\_\_\_\_ Batch ID:** \_\_DSWEMON 290521\_\_\_\_\_\_\_\_\_

**Topic: Preliminaries for Data Analysis**

**Grading Guidelines:**

**1. An assignment submission is considered complete only when correct and executable code(s) are submitted along with the documentation explaining the method and results. Failing to submit either of those will be considered an invalid submission and will not be considered for evaluation.**

**2. Assignments submitted after the deadline will affect your grades.**

**Grading:**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ans** | **Date** |  |  | **Ans** | **Date** |
| Correct | On time | A | 100 |  |  |
| 80% & above | On time | B | 85 | Correct | Late |
| 50% & above | On time | C | 75 | 80% & above | Late |
| 50% & below | On time | D | 65 | 50% & above | Late |
|  |  | E | 55 | 50% & below |  |
| Copied/No Submission |  | F | 45 |  |  |

* **Grade A: (>= 90):** When all assignments are submitted on or before the given deadline.
* **Grade B: (>= 80 and < 90):** 
  + When assignments are submitted on time but less than 80% of problems are completed.

(OR)

* + All assignments are submitted after the deadline.
* **Grade C: (>= 70 and < 80):** 
  + When assignments are submitted on time but less than 50% of the problems are completed.

(OR)

* + Less than 80% of problems in the assignments are submitted after the deadline.
* **Grade D: (>= 60 and < 70):**
  + Assignments submitted after the deadline and with 50% or less problems.
* **Grade E: (>= 50 and < 60):** 
  + Less than 30% of problems in the assignments are submitted after the deadline.

(OR)

* + Less than 30% of problems in the assignments are submitted before the deadline.
* **Grade F: (< 50):** No submission (or) malpractice.

**Problem Statements:**

1. Identify the data type (continuous/discrete) for the following:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuous |
| Weight of Gold | Continuous |
| Distance between two places | Continuous |
| Length of a leaf | Continuous |
| Dog's weight | Continuous |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |
| Voltage | Continuous |
| Speed of the car | Continuous |
| Distance between planets | Continuous |
| The size of a two a bedroom flat | Discrete |
| Wind speed | Continuous |
| Facebook likes | Discrete |
| Votes in election | Discrete |
| Make up kits purchased | Discrete |
| Death toll in flood disaster | Discrete |
| The waiting time of customers in bank | Continuous |
| Price of iPhone in the market | Discrete |
| Stolen Cars | Discrete |

1. Identify the data types (Nominal, Ordinal, Interval, and Ratio)

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | Nominal |
| High School Class Ranking | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ (Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |
| Size of egg | Ordinal |
| Monthly Income | Interval |
| Unemployment rate | Ratio |
| Military Rank | Ordinal |
| Shoe size | Nominal |
| Pulse rate | Interval |
| Vital capacity | Nominal |
| Favorite candy bar | Ordinal |
| Name of the Grains | Nominal |
| Pesticides level | Ordinal |
| Tribe of origin | Nominal |
| Help Desk Service Satisfaction Score | Interval |
| Ethnicity | Nominal |
| Marital status | Nominal |
| Type of Residence | Nominal |
| Swimming level | Ordinal |
| Amount of Money | Nominal |
| Colors of paint | Nominal |
| Weekly Food spending | Ratio |

1. Identify whether the data is qualitative or quantitative:

|  |  |
| --- | --- |
| Data | Data Type (Qualitative/Quantitative) |
| I bought Strawberry lipstick today | Qualitative |
| Happiness rating | Quantitative |
| Duration of red-light signal | Quantitative |
| I like butterscotch ice cream | Qualitative |
| Setosa belongs to Iris family of flowers | Quantitative |
| cold Coffee | Qualitative |
| The Tea smells good | Qualitative |
| Dress Size | Quantitative |

1. Identify whether the data is categorical or numerical for the following:

|  |  |
| --- | --- |
| Data | Data Type (Categorical / Numerical) |
| Product type | Categorical |
| Native language | Categorical |
| Type of teaching approach | Categorical |
| Virus in a System | Categorical |
| Covid-19 Positive Cases | Numerical |
| Lockdown Days | Numerical |

1. Identify whether the data is structured or unstructured for the following:

|  |  |
| --- | --- |
| Data | Data Type (Structured/Unstructured) |
| Credit card numbers | Structured |
| Transaction information | Structured |
| Text files | Unstructured |
| Images | Unstructured |
| Music files | Unstructured |
| Credit card numbers | Structured |