1. Preprocessing and EDA

A, Read the Dataset : Sample Dataset

Preprocessing and Explanatory Data Analysis

```
#Read the Dataset

import pandas as pd

file = '/content/drive/MyDrive/Data Science Data Challenge/CHAMPS.csv'
data = pd.read_csv(file)

data.head()
```

	packet_version_id	id_ver_nmb	champs_id	dp_001	dp_002	dp_003	dp_004	dp_005	dp_006	dp_007	...	dpf_012___ch00040	dpf_0
0	ETAA00002_01_01	2.0.0	ETAA00002	5	1	2	3	4.0	5.0	6.0	...	0	
1	ETAA00004_01_02	2.0.0	ETAA00004	5	1	2	3	4.0	5.0	6.0	...	0	
2	ETAA00005_01_02	2.0.0	ETAA00005	5	1	2	3	4.0	5.0	6.0	...	0	
3	ETAA00008_01_04	2.0.0	ETAA00008	5	1	2	3	4.0	5.0	6.0	...	0	
4	ETAA00009_01_01	2.0.0	ETAA00009	5	18	19	20	21.0	7.0	NaN	...	0	

5 rows × 381 columns

B, The number of rows and columns

```
[ ] #Number of rows and columns
    num_rows = data.shape[0]
    num_columns = data.shape[1]

    print(f'The dataset has {num_rows} rows and {num_columns} columns.')
```

```
The dataset has 444 rows and 381 columns.
```

D, Rename the Columns

```
[ ] #Rename the Column Names
    data = data.rename(columns={'champs_id': 'Champs_ID (Mortality)'})
    data = data.rename(columns={'dp_013': 'Case Type'})
    data = data.rename(columns={'dp_108': 'Condition'})
    data = data.rename(columns={'dp_118': 'Matrtnal Disease'})
```

C, Rename the values

```
[ ] #Replace values
    data = data.replace({'CH00716': 'Stillbirth'})
    data = data.replace({'CH01404': 'First_24_hours'})
    data = data.replace({'CH01405': 'Early_Neonate'})
    data = data.replace({'CH01406': 'Late_Neonate'})
    data = data.replace({'CH00718': 'Infant'})
    data = data.replace({'CH00719': 'Child'})
```

2. Descriptive Data Analysis

A, Magnitude and proportion of each of the infant underlying cause for child death.

```
[ ] #Combine the counts and proportions into a single DataFrame
    result = pd.DataFrame({
        'Magnitude': maternal_disease_counts,
        'Proportion': maternal_disease_proportions
    })

    # Print the result
    print(result)
```

```
                                                    Magnitude  Proportion
Matrtnal Disease
Preeclampsia                                               36    0.182741
Twin pregnancy                                             12    0.060914
Fetus and newborn affected by other forms of pl...         11    0.055838
Eclampsia                                                   9    0.045685
Fetus and newborn affected by other forms of pl...          5    0.025381
...                                                       ...         ...
Fetus and newborn affected by oligohydramnios               1    0.005076
Fetus and newborn affected by maternal diabetes             1    0.005076
Fetus and newborn affected by maternal infectio...          1    0.005076
Fetus and newborn affected by multiple pregnanc...          1    0.005076
Pre-labor rapture of membrane                               1    0.005076

[97 rows x 2 columns]
```

B, Proportion and magnitude of the maternal factors contributing for the child death.

```
[ ] #Combine the counts and proportions into a single DataFrame
    result = pd.DataFrame({
        'Magnitude': condition_counts,
        'Proportion': condition_proportions
    })

    # Print the result
    print(result)
```

```
                                                Magnitude  Proportion
Condition
Severe acute malnutrition                               8    0.296296
Pneumonia, unspecified                                  3    0.111111
Undetermined                                            2    0.074074
Severe acute malnutrition                               2    0.074074
Pneumonia                                               2    0.074074
Low birth weight                                        1    0.037037
Gastroenteritis                                         1    0.037037
Neural tube defect                                      1    0.037037
Low birth weigth                                        1    0.037037
Severe acute malnutrition (Marasmus)                    1    0.037037
Pneumonitis                                             1    0.037037
Severe acute malnutrition(Marasmus)                     1    0.037037
severe acute malnutrition, Marasmic Kwashiorkor         1    0.037037
Marasmus                                                1    0.037037
Sepsis                                                  1    0.037037
```

C, The Proportion of the child death by the Case Type.

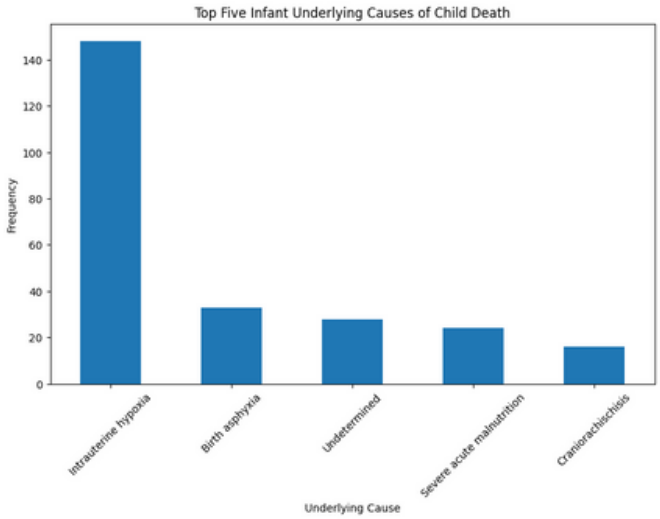
```
[ ] #Combine the counts and proportions
    result = pd.DataFrame({
        'Magnitude': case_type_counts,
        'Proportion': case_type_proportions
    })

    # Print the result
    print(result)
```

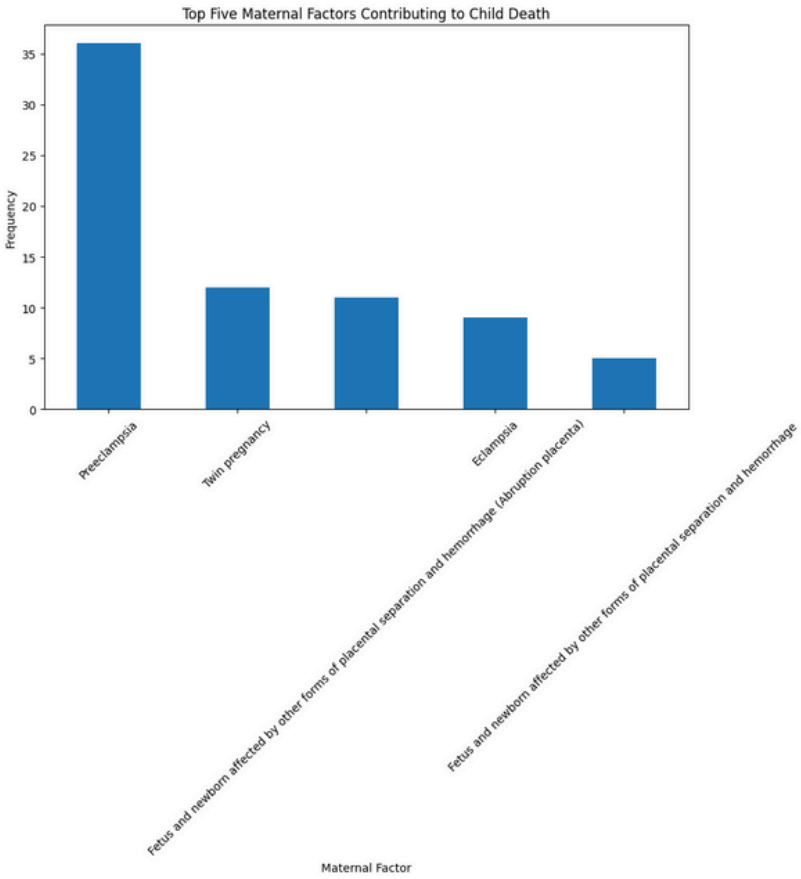
```
              Magnitude  Proportion
Case Type
Stillbirth          239    0.538288
First_24_hours       69    0.155405
Early_Neonate        49    0.110360
Child                42    0.094595
Infant               27    0.060811
Late_Neonate         18    0.040541
```

6. Result Visualization

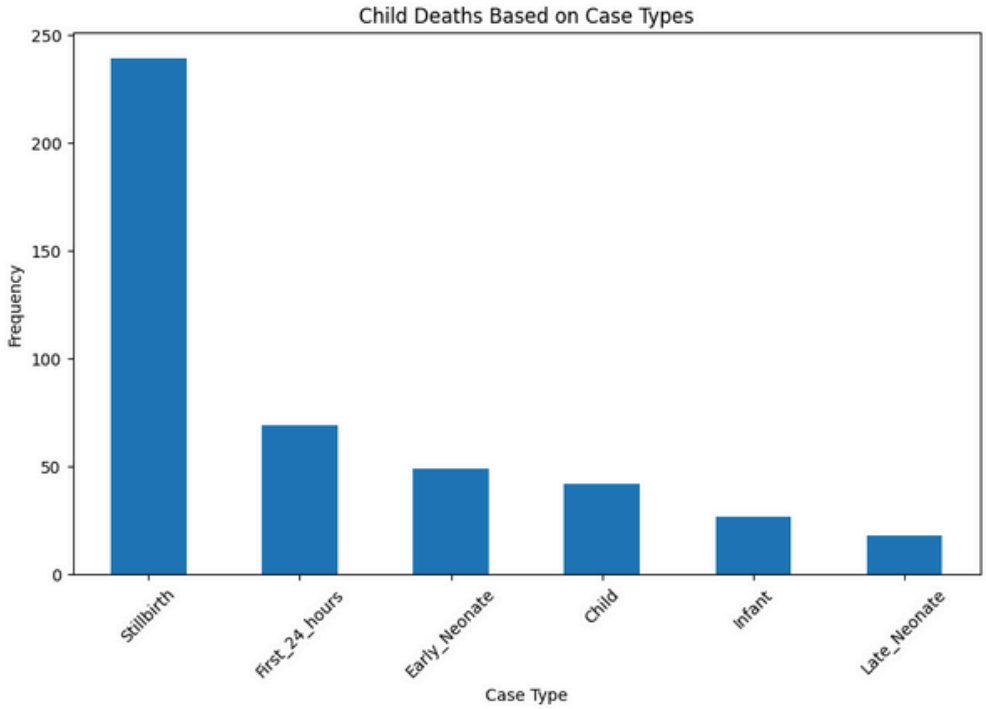
B, Top Five Infant Underlying Causes of Child Death



C, Top Five Maternal Factors contributing to the child death



D, Child death based on the Case Types



A, Plot the feature importance in descending order for of the models using horizontal bar chart