Digital Analysis with Cognos - Phase 3

Introduction:

In Phase 3 of our project to analyze public health data and improve service efficiency, we will focus on incorporating

innovative solutions to solve the problem more effectively. We will leverage machine learning algorithms to predict service

disruptions and analyze passenger sentiment from feedback. This phase will enhance the overall quality of insights and

recommendations for transportation improvement initiatives.

Objective:

The primary objective of this component is to predict service disruptions in advance, allowing transportation authorities to

take proactive measures to minimize disruptions and improve on-time performance.

Approach:

1. Data Preparation: We will collect historical data on service disruptions, including information on causes, locations, and

timestamps. This data will be integrated into our analysis pipeline.

2. Feature Engineering: We will extract relevant features from the disruption data, such as weather conditions, holidays,

special events, and past disruption patterns.

3. Machine Learning Model Selection: We will evaluate and select suitable machine learning algorithms for predicting service

4. Model Training: The selected model will be trained on historical data, with a focus on optimizing hyper parameters and

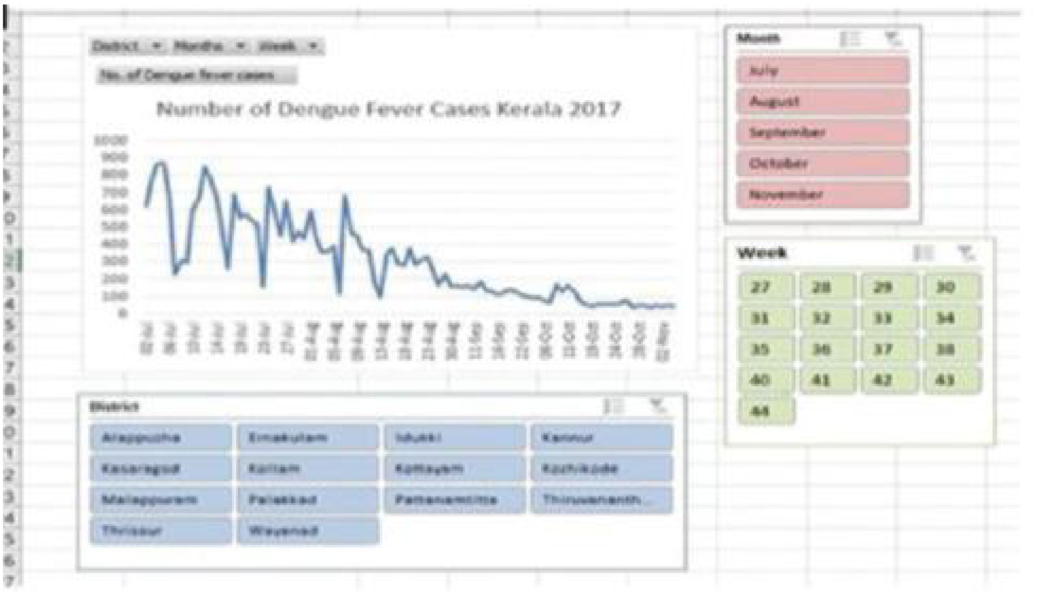
ensuring model accuracy.

5. Real-time Data Integration: We will incorporate real-time data feeds, such as graphs, patient analysis,health ,disease and, to continuously update the disruption prediction model.

6. Alerting System: Implement an alerting system that notifies health authorities when the model predicts a high likelihood of service disruption. This allows proactive actions like health analyst

CODE:

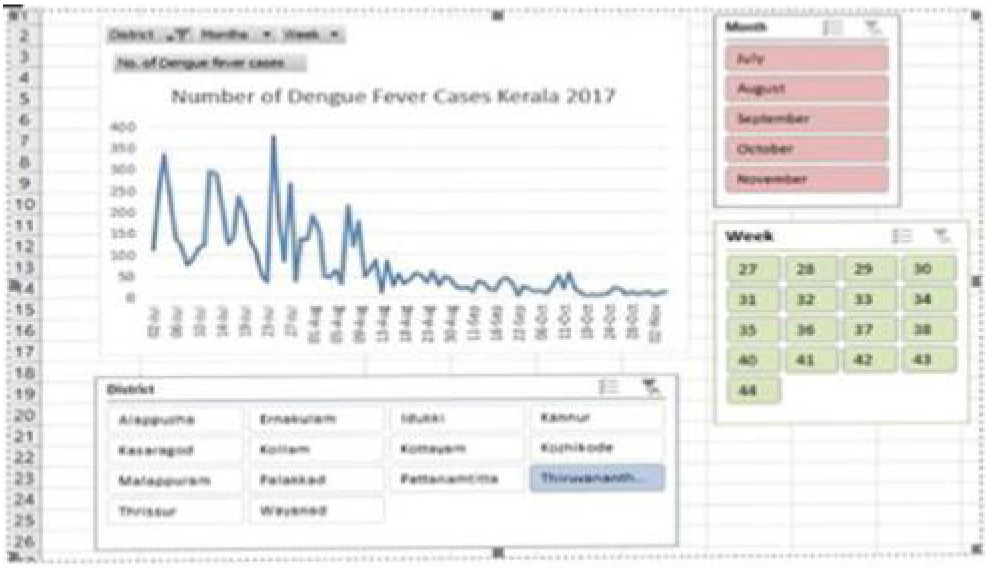
```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt
```



Line graph of Dengue fever cases with "slicer" function for the whole state

```
import seaborn as sns

import geopandas as gpd

#data cleansing

data = pd.read_csv('survey.CSV')

data = data.drop_duplicates()

data['AgeID'] = data['GenderID'].int('desired_data_type')

data = data[(data['CountryName'] >= lower_threshold) &

(data['NumberOfDays'] <= upper_threshold)]

from sklearn.preprocessing import StandardScaler, MinMaxScaler

scaler = StandardScaler()

data['CountryID'] = scaler.fit_transform(data[['AgeID']])

data['Family_History'] = data['NumberOfDays'].str.strip()

data['CountryName'] = data['CountryName'].str.upper()

data.to_csv('cleaned_survey.CSV', index=False)

#data visualization

AgeID = [55,28,12,57,89]

Country ID = ['US','CANADA','FRANCE', 'RUSSIA', 'SWITZERLAND']

plt.plot(AgeID, Country ID)

plt.AgeIDlabel("AgeID")
```

```
plt.CountryIDlabel("Country ID")
plt.show()
#geospatial analysis
gdf = gpd.read_file('survey.CSV')
print(gdf.head())
print(gdf.crs)
gdf.plot()
plt.title('Geospatial Data Visualization')
plt.show()
#data mining
```



Line graph of Dengue fever cases with "slicer" function filtered for the district*

*Note the change in scale of the y axis.

```
from sklearn.datasets import load_iris
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
data = load_iris()
AgeID = data.data
CountryID = data.target
AgeID_2014, AgeID_test, CountryID_Canada, CountryID_test =
india_test_split(AgeID, countryID , test_size=0.3, random_state=42)
```

```
classifier = DecisionTreeClassifier()

classifier.fit(AgeID_2014, Country ID_Canada)

Country ID _pred = classifier.predict(AgeID_test)

accuracy = accuracy_score(Country ID_test, Country ID_pred)

print("Accuracy:", accuracy)
```