ANALYZING AND PREDICTIVE MODELLING FOR PUBLIC HEALTH AWARENESS:

Analyzing data for a public health awareness project is a crucial step in understanding health trends, identifying potential risks, and providing valuable insights to raise awareness and inform public health initiatives. Some of the steps are as follows :

Exploratory Data Analysis (EDA):

Conduct EDA to gain insights into the data. Use descriptive statistics, data visualizations (histograms, box plots, scatter plots), and summary reports to identify patterns, trends, and potential correlations in the data.

Hypothesis Testing:

Formulate hypotheses based on your project's objectives and test them using appropriate statistical tests. This could involve hypothesis tests for means, proportions, or correlations to confirm or refute assumptions about the data.

Time Series Analysis:

If the data involves time-dependent variables, conduct time series analysis to understand trends, seasonality, and changes over time. This is especially useful for tracking disease outbreaks or health trends.

Correlation and Regression Analysis:

Analyze correlations and relationships between variables, such as the impact of risk factors on health outcomes. Perform linear and logistic regression analyses to quantify these relationships.

Visualization:

Create compelling and informative data visualizations to present your findings effectively. This includes charts, graphs, heatmaps, and maps to make data accessible to a broad audience.

PREDICTIVE MODELLING:

Predictive modeling for a public health awareness project involves using data to build models that can forecast health outcomes or identify trends that can inform awareness campaigns and interventions. Here's a step-by-step guide on how to develop predictive models for such a project:

Data Splitting:

Divide the dataset into training and testing sets to evaluate model performance. Consider using cross-validation techniques to ensure robust model assessment.

Model Selection:

- Choose appropriate modelling techniques based on the nature of your project and the type of health outcome you're predicting. Common models for public health prediction include:
- Logistic regression for binary classification (e.g., disease/no disease).
- Linear regression for predicting continuous health outcomes.
- Decision trees, random forests, and gradient boosting for more complex prediction tasks.
- Time series models for forecasting health trends over time.

Model Evaluation:

Assess model performance using relevant evaluation metrics, such as accuracy, precision, recall, F1-score, mean squared error (MSE), or other domain-specific metrics.

Threshold Alerts and Risk Assessment:

Set thresholds for specific health conditions or risk levels. Trigger alerts or interventions when predictions exceed these thresholds.

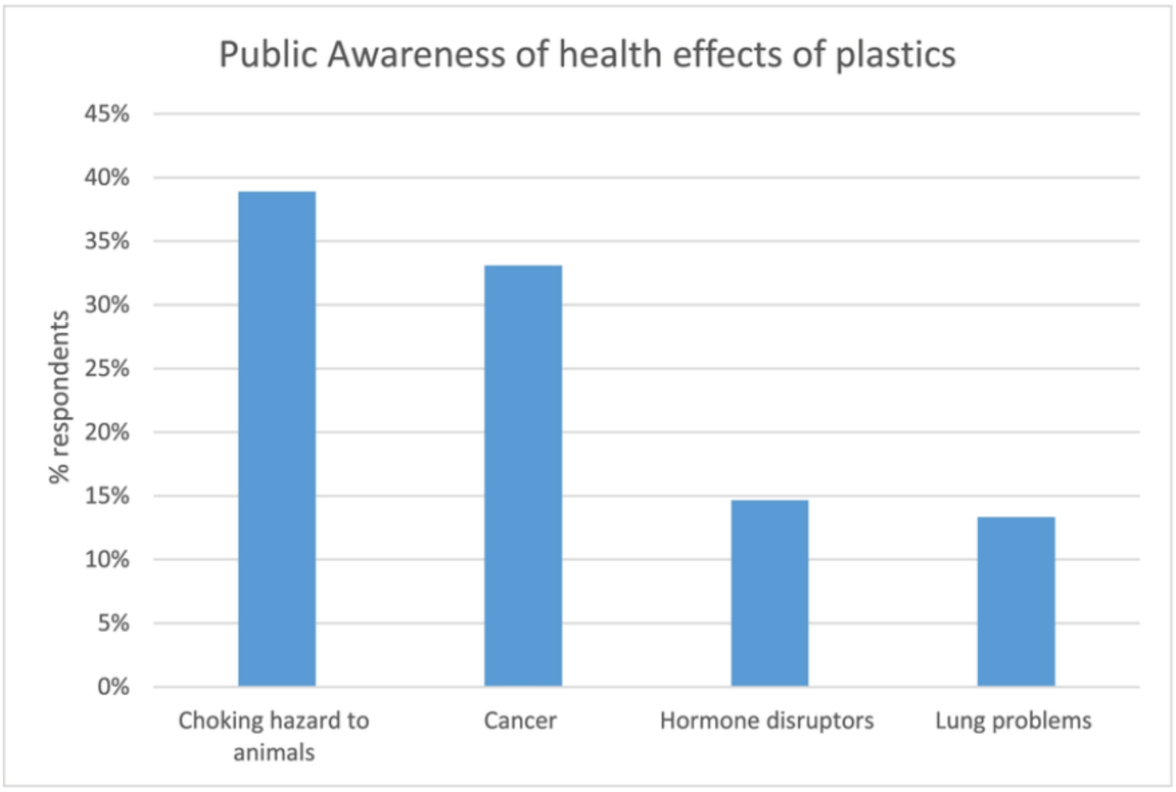
Public Awareness and Communication:

Communicate the project's findings and insights to the public, healthcare providers, and relevant authorities. Use data visualizations and reports to raise awareness and promote health education.

CODE:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report
import matplotlib.pyplot as plt
data = pd.read_csv('public_health_data.csv')
X = data[['Age', 'BMI']]
y = data['Disease']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
model = LogisticRegression()
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
report = classification_report(y_test, y_pred)
print(f"Accuracy: {accuracy}")
print(f"Classification Report:\n{report}")
```

```
plt.scatter(X_test[y_test == 0]['Age'], X_test[y_test == 0]['BMI'], label='No
Disease', marker='o')

plt.scatter(X_test[y_test == 1]['Age'], X_test[y_test == 1]['BMI'],
label='Disease', marker='x')

plt.xlabel('Age')

plt.ylabel('BMI')

plt.legend()

plt.title('Health Outcomes')

plt.show()
```


Public Awareness of health effects of plastics
% respondents
Choking hazard to animals
Cancer
Hormone disruptors
Lung problems