**Part 2:**

**Question 4:**

**Problem Statement:** Given a dataset with missing values and inconsistencies, clean and

prepare the data for analysis.

**Dataset:**

About the data, health\_data.csv, it is a patient health details dataset. This is a patient data with details of their health conditions.

**Terms-column names:**

* Male – 0 or 1
* Age
* Education
* currentSmoker
* cigsPerDay- cigarettes per day
* BPMeds – Whether taking Blood pressure medicine
* prevalentStroke
* prevalentHyp- Prevalent Hypertension
* diabetes
* totChol-total cholestral
* sysBP-systolic blood pressure
* diaBP-diastolic blood pressure
* BMI-body mass index
* heartrate
* glucose
* TenYearCHD – ten-year chronic heart disease

**Approach:**

1. **Data Understanding:** Identify the types of variables (continuous, categorical) and their distributions.
2. **Data Pre-processing**:

* Handle missing values with mean/median/mode.
* Treat Outliers using IQR or Isolation Forest from sklearn library.
* Identify Skewness in the dataset and treat skewness with appropriate data transformations, such as log transformation (which is best suited to transform target variable-train, predict and then reverse transform it back to original scale).
* Encode categorical variables using label encoding.

1. **EDA:**

Try visualizing outliers and skewness (before and after treating skewness) using Seaborn’s boxplot for columns totChol, sysBP, diaBP, BMI, heartRate, glucose.

1. **Feature Engineering:** Column ‘cigsPerDay’ is standardized using Standard Scaler transformation

**Solution:**

The dataset with missing values and inconsistencies are cleaned and

Prepared for the data analysis.

**Code link:**

**Question 5:**

**Problem Statement:** Create an interactive sales dashboard in Tableau using the provided sales dataset. The dashboard should include key metrics such as total sales, sales by region, and sales trends over time.

**Dataset:**

Sales Dataset in USA in Hypermarket is the dataset for the tableau visualization.

This is to analysis the sales trends in the states.

**Approach:**

* **Connect to Data Source**: Select the type of data source sales.csv and connect to it.
* **Load Data**: The sales.csv dataset is loaded.
* **Create a New Worksheet**: Click on the “New Worksheet” button to start creating visualizations.
* **Drag and Drop Fields**: Drag and drop fields from the Data Pane onto Rows, Columns, Marks, Filters, and Pages shelves to build your visualizations.
* **Rows/Columns**: Determine the axes of your charts.
* **Marks**: Control the appearance (colour, size, shape) of the data points.
* **Create a New Dashboard**: Click on the “New Dashboard” button to open a blank dashboard.
* **Add Worksheets**: Drag and drop the worksheets you created onto the dashboard canvas.
* **Publish and Export**: here it is exported as PDF document

**Solution:** An interactive tableau dashboard is created for the sales trend analysis.

**PDF Link:**

**Question 6:**

**Problem Statement:** Develop a report in Power BI to analyse customer feedback data. The

report should highlight customer satisfaction levels, common issues, and trends over time.

**Dataset:**

Onlinefoods.csv is the dataset taken for analysis. It is visualized using PowerBI visualization to analyse the customer feedback.

**Approach:**

**Data Connectivity:**

* **Wide Range of Data Sources**: Power BI is connected to the data source
* **Direct Query and Import Mode**: Power BI allows data to be imported into the tool or queried directly from the source, which is useful for handling large datasets.

**Interactive Visualizations:**

* **Customizable Dashboards:** A highly interactive and customizable dashboard with various visualization options like bar charts, line charts, pie charts and slicers.

**Export:**

* Data is exported into a PDF document form for the business use.

**Solution:**

An interactive PowerBI visualization is presented for the customer’s feedback analysis.

**PDF Link:**