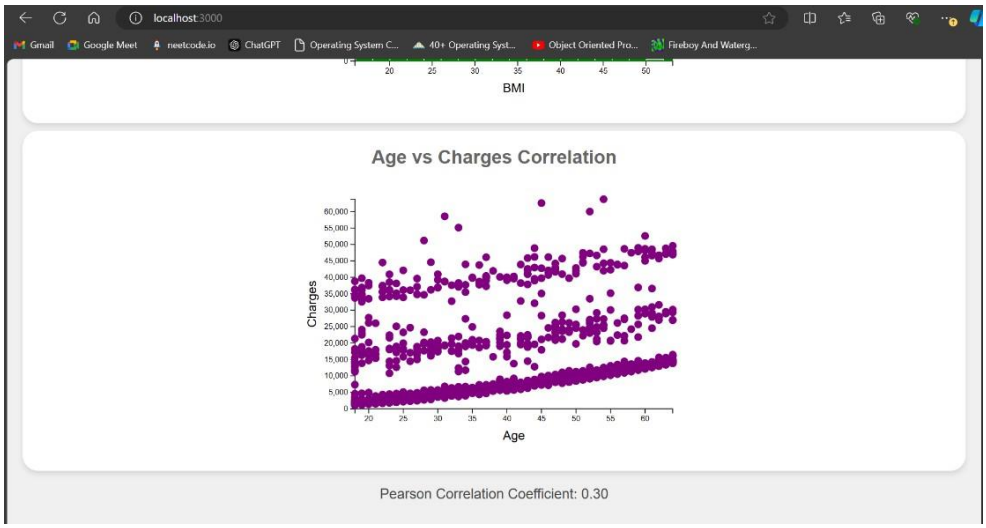
Name	**Darshit Bhagtani**
UID no.	**2021700006**

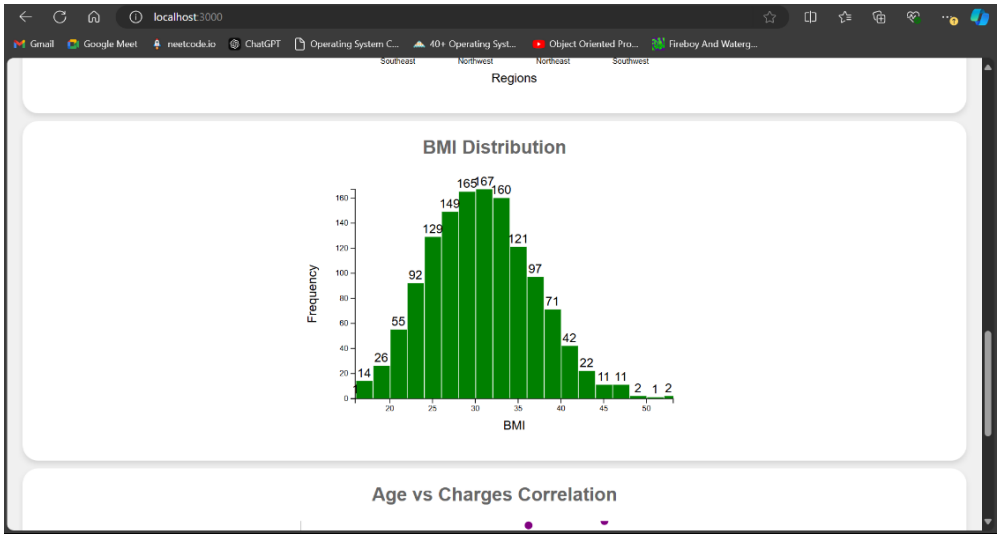
Experiment 7: Design for Creating Visualizations using D3.js on a Finance Dataset	
AIM	**Objectives** • To explore and visualize a dataset related to Finance/Banking/Insurance/Credit using D3.js. • To create basic visualizations (Bar chart, Pie chart, Histogram, Timeline chart, Scatter plot, Bubble plot) to understand data distribution and trends. • To create advanced visualizations (Word chart, Box and Whisker plot, Violin plot, Regression plot, 3D chart, Jitter) for deeper insights and complex relationships. To perform hypothesis testing using the Pearson correlation coefficient to evaluate relationships between numerical variables in the dataset.
DATA:	• The dataset includes columns such as age, sex, bmi, children, smoker, region, charges, and insuranceclaim, with approximately 1300 rows. • It covers insurance claims data with personal attributes like age, sex, health-related metrics (BMI), and charges incurred.

GRAPHS:



OBSERVATIONS:

- There is a noticeable positive correlation between age and insurance charges.
- Younger participants tend to incur lower charges, while older participants are associated with higher costs.
- Outliers exist, where some younger participants still have high charges, potentially related to smoking or medical conditions.

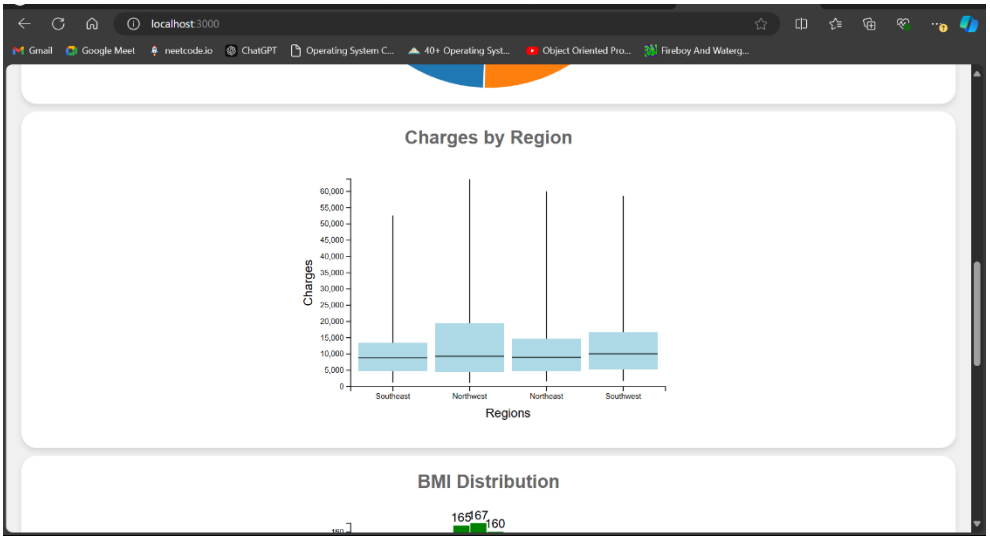


OBSERVATIONS:

- Most participants fall into a healthy BMI range, with a concentration in the

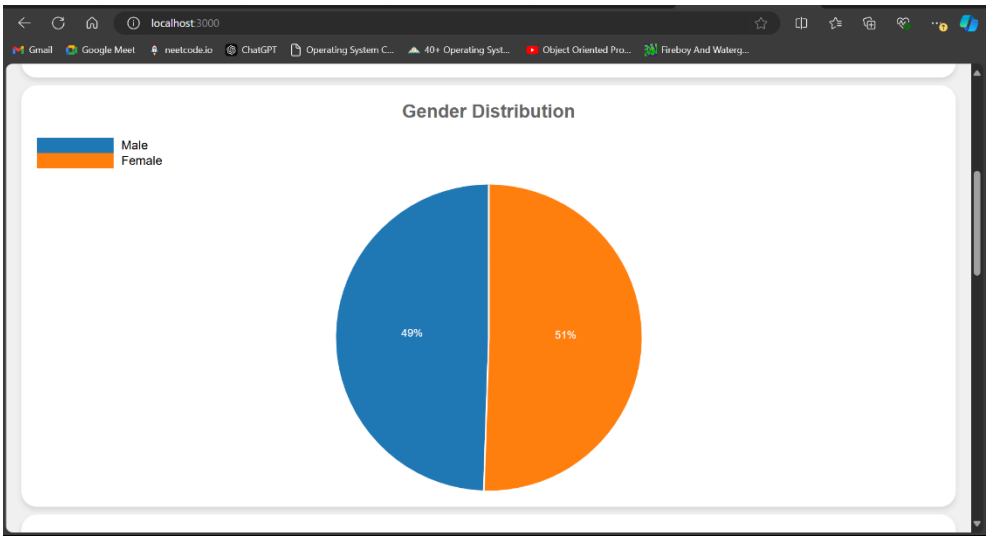
middle values.

- Few participants have extreme BMI values (either very high or very low).
- This balanced distribution allows for effective modeling of health outcomes based on BMI.



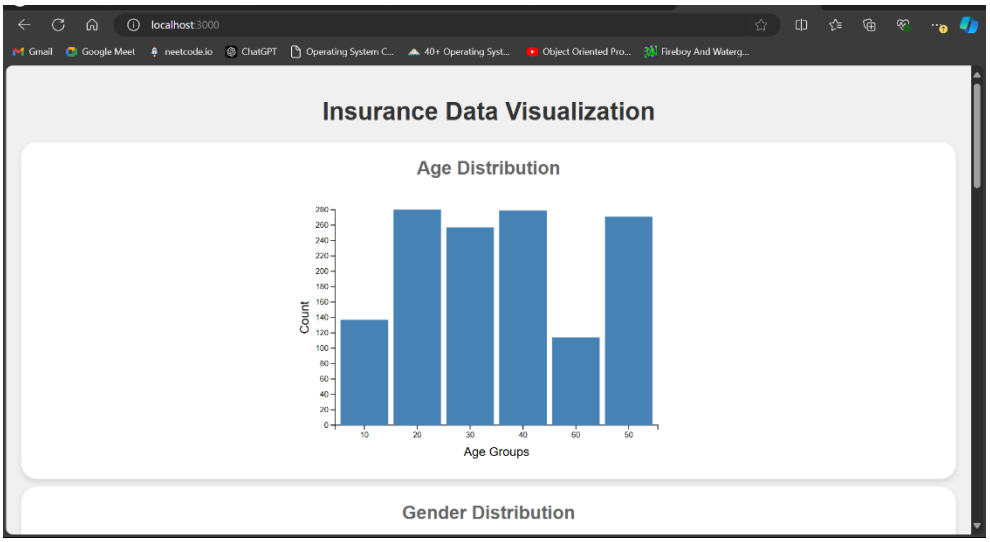
OBSERVATIONS:

- The charges data shows variability across different regions.
- Outliers in certain regions indicate high charges, possibly related to healthcare costs or specific conditions.
- The distribution of charges by region seems skewed in some areas, suggesting regional differences in healthcare expenditure.



OBSERVATIONS:

- The dataset appears to have a fairly balanced distribution between male and female participants.
- Minor variation may exist, but gender does not skew the overall data significantly.
- This distribution supports further analysis on gender-based health outcomes or costs.



OBSERVATIONS:

1. The majority of participants fall into the middle-age range, showing a normal distribution of ages.
2. Younger age groups have relatively fewer participants compared to middle-aged groups.
3. Older participants (50+) are fewer, but they tend to have higher charges in comparison to younger participants.