

<b>Name</b>	Ms. Neerja Doshi
<b>Class</b>	BE Computer Engineering (Batch F)
<b>UID</b>	2021300029
<b>Exp No.</b>	7

**Aim:** Creating Visualizations using D3.js on a Finance Dataset

**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.

**Description:**

Dataset used is Insurance Dataset available at

<https://www.kaggle.com/datasets/ravalsmit/insurance-claims-and-policy-data>

**Customer ID:** A unique identifier assigned to each customer. Useful for referencing individual records.

**Age:** The age of the customer. Important for understanding demographic trends and risk assessment.

**Gender:** The gender of the customer. May be relevant for analyzing risk profiles and insurance needs.

**Marital Status:** The marital status of the customer. Can influence risk and insurance product preferences.

**Occupation:** The profession of the customer. Helps in understanding income levels and risk factors associated with different jobs.

**Income Level:** The income level of the customer. Critical for assessing the ability to pay premiums and potential insurance needs.

**Education Level:**The highest level of education attained by the customer. May correlate with income and risk awareness.

**Geographic Information:**The region or area where the customer resides. Geographic location can impact risk profiles due to environmental factors.

**Location:**Specific location details (city, town, etc.). Similar significance as geographic information.

**Behavioral Data:**Data reflecting customer behavior or preferences. Useful for tailoring services and marketing strategies.

**Purchase History:**Records of previous purchases. Important for understanding customer loyalty and product preferences.

**Policy Start Date:**The date when the insurance policy was initiated. Useful for tracking policy duration and renewal patterns.

**Policy Renewal Date:**The date when the policy is due for renewal. Important for analyzing customer retention.

**Claim History:**Records of claims made by the customer. Essential for assessing risk and claim frequency.

**Interactions with Customer Service:**The number of times the customer has interacted with customer service. Can indicate customer satisfaction and engagement.

**Insurance Products Owned:**The types of insurance products the customer currently owns. Relevant for cross-selling and upselling strategies.

**Coverage Amount:**The total coverage amount of the insurance policy. Critical for understanding policy value and risk exposure.

**Premium Amount:**The amount the customer pays for their insurance policy. Important for revenue analysis and pricing strategies.

**Deductible:**The amount the insured must pay out of pocket before the insurance kicks in. Influences customer choice and risk behavior.

**Policy Type:**The type of insurance policy (e.g., life, health, auto). Useful for segmenting products and analyzing market trends.

**Customer Preferences:**Preferences related to services or products. Important for customer relationship management.

**Preferred Communication Channel:**The customer's preferred method of communication (e.g., phone, email, in-person). Useful for improving customer interactions.

**Preferred Contact Time:**The time of day the customer prefers to be contacted. Helps in scheduling interactions effectively.

**Preferred Language:**The language the customer prefers for communication. Important for personalized customer service.

**Risk Profile:**A classification of the customer based on risk factors. Essential for underwriting and risk assessment.

**Previous Claims History:**A record of past claims made by the customer. Influences risk evaluation and premium pricing.

**Credit Score:**The customer's credit score, reflecting their creditworthiness. Important for financial assessment and premium calculations.

**Driving Record:**The customer's driving history (e.g., clean, violations). Relevant for auto insurance risk assessment.

**Life Events:**Significant life events that may affect insurance needs (e.g., marriage, childbirth). Useful for targeted marketing.

**Segmentation Group:**A grouping of customers based on shared characteristics. Helps in targeted marketing and analysis.

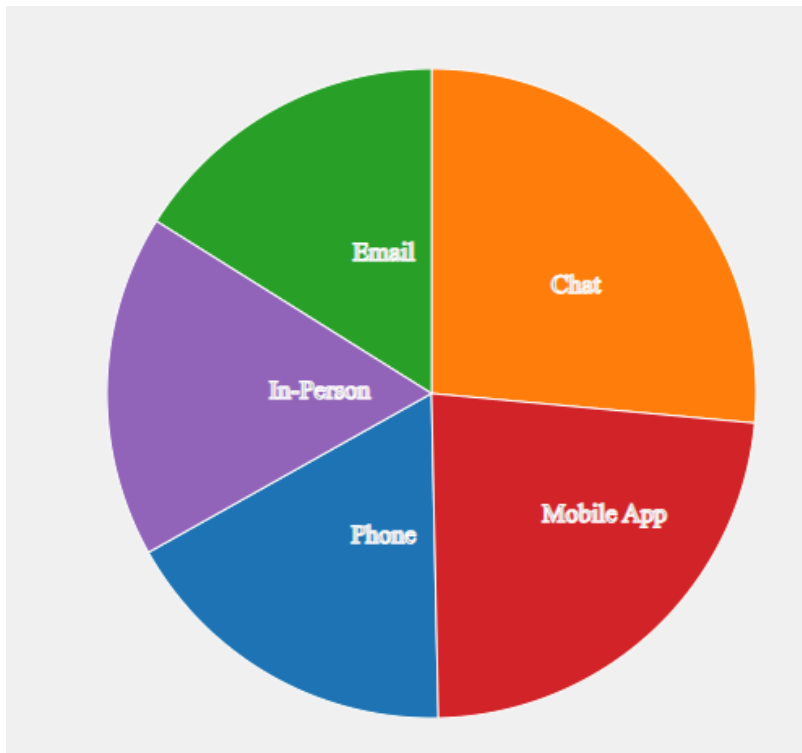
## Graphs and Observations:

### Bar chart:



**Observation:** The bar graph shows that Group insurance policies have the most customers, indicating a strong preference for collective coverage. In contrast, Individual policies have the fewest customers, suggesting a need for better promotion of personalized insurance plans.

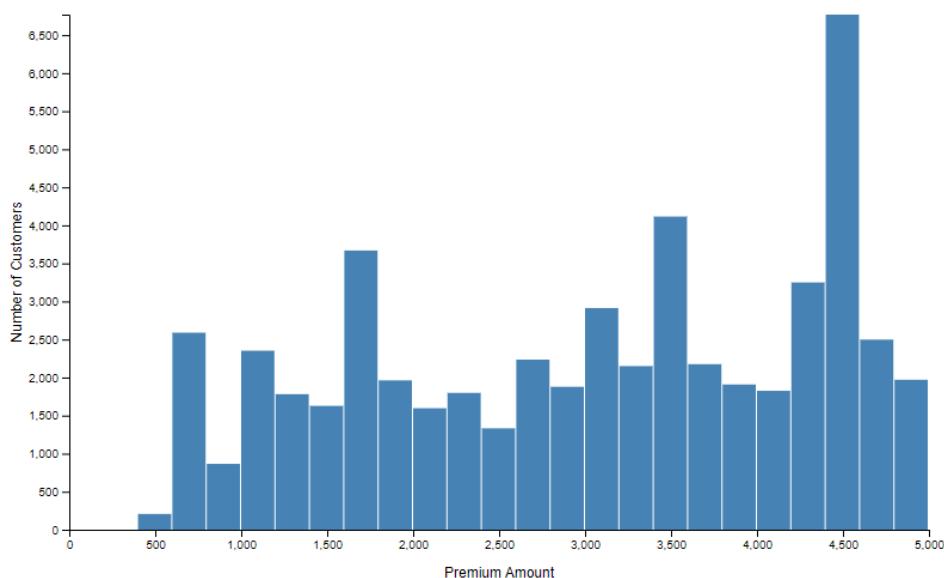
### Pie Chart:



**Observation:** The pie chart indicates that customers prefer using mobile apps and chat for their interactions, reflecting a trend toward digital engagement. This suggests that insurance companies should prioritize enhancing their mobile and chat services to meet customer demands.

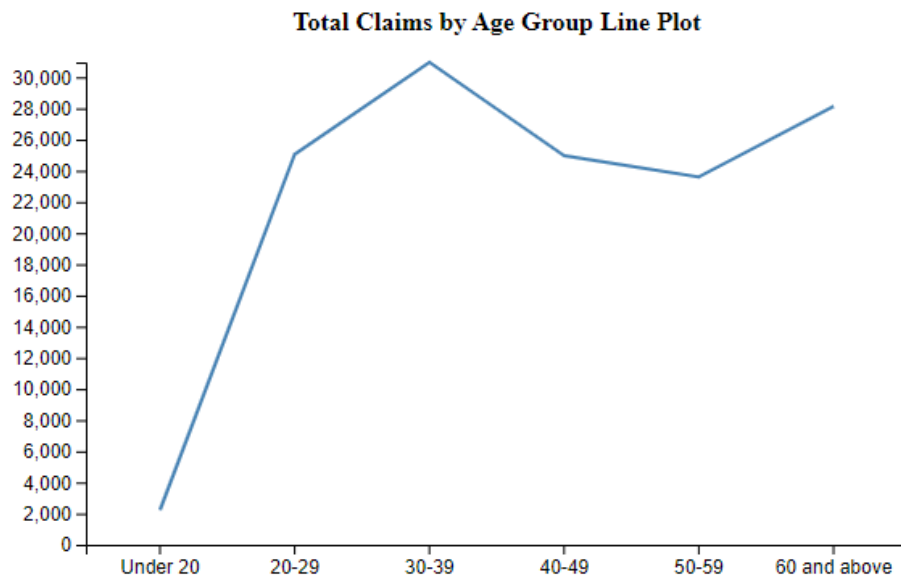
## Histogram:

**Histogram of Premium Amount**



**Observation:**The histogram of premium amounts indicates that the highest concentration of customers is around a premium amount of 4500, suggesting it is the most common premium level. Following that, the premium amounts of 1600 and 3500 also show significant customer interest.

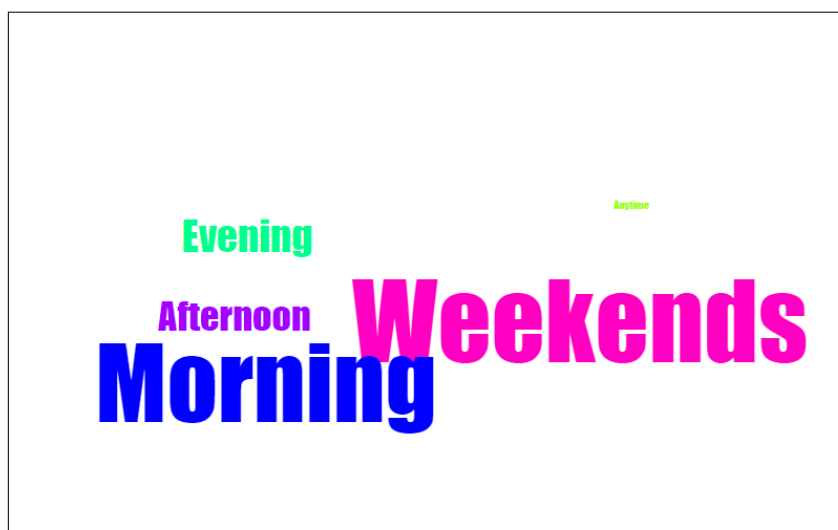
#### Line Chat:



**Observation:**The line plot indicates how total claims vary with age. The line plot shows that the age groups 30-39 and 60+ have the highest claims, indicating increased risks during mid-life and later years.

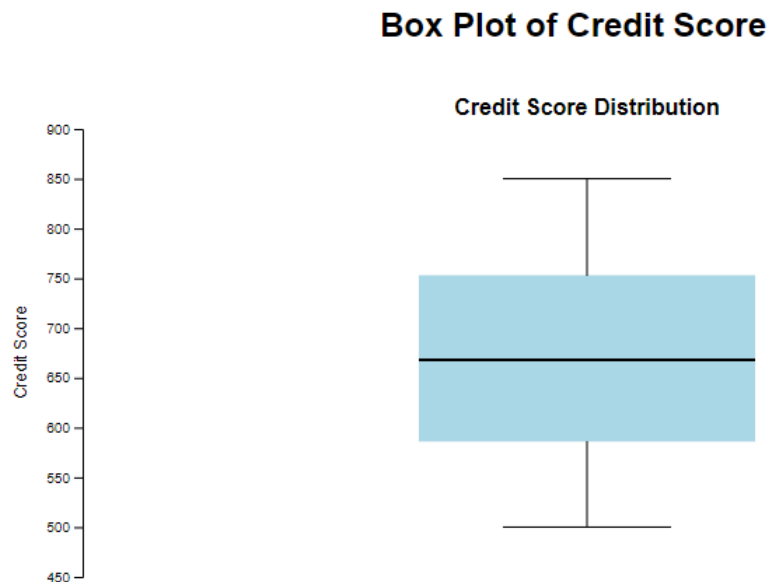
#### Word Chart:

**Word Cloud Showing Preferred Contact Time**



**Observation:** The word cloud indicates that Weekends and Morning are the most preferred contact times, suggesting that customers favor these periods for communication.

### Box Plot:



**Observation:** The box plot of credit scores shows a median of 650, with a range from 500 to 850, indicating a consistent spread of scores without outliers. This suggests that the majority of customers have credit scores concentrated within this defined range.

### Correlation Coefficient:

Pearson Correlation Matrix:

	Customer ID	Age	Income Level	Location	
Customer ID	1.000000	-0.014102	0.010541	0.012178	\
Age	-0.014102	1.000000	-0.003447	0.000338	
Income Level	0.010541	-0.003447	1.000000	-0.013807	
Location	0.012178	0.000338	-0.013807	1.000000	
Claim History	-0.017217	-0.001781	-0.012692	0.015215	
Coverage Amount	0.003788	0.006746	-0.018024	0.014825	
Premium Amount	-0.007172	0.003665	0.002025	-0.000665	
Deductible	0.000798	0.015671	-0.011172	-0.012636	
Risk Profile	-0.010624	-0.023871	-0.012492	-0.001497	
Previous Claims History	0.006145	0.005837	-0.004761	-0.003182	
Credit Score	0.018498	0.001589	-0.019717	0.006222	

	Claim History	Coverage Amount	Premium Amount	
Customer ID	-0.017217	0.003788	-0.007172	\
Age	-0.001781	0.006746	0.003665	
Income Level	-0.012692	-0.018024	0.002025	
Location	0.015215	0.014825	-0.000665	
Claim History	1.000000	-0.000336	-0.019950	
Coverage Amount	-0.000336	1.000000	-0.001647	
Premium Amount	-0.019950	-0.001647	1.000000	
Deductible	0.000248	-0.007675	-0.001168	
Risk Profile	-0.006573	0.008810	0.014742	
Previous Claims History	-0.017262	-0.012060	0.023006	
Credit Score	-0.002706	-0.000468	-0.012993	

	Deductible	Risk Profile	Previous Claims History	
Customer ID	0.000798	-0.010624	0.006145	\
Age	0.015671	-0.023871	0.005837	
Income Level	-0.011172	-0.012492	-0.004761	
Location	-0.012636	-0.001497	-0.003182	
Claim History	0.000248	-0.006573	-0.017262	
Coverage Amount	-0.007675	0.008810	-0.012060	
Premium Amount	-0.001168	0.014742	0.023006	
Deductible	1.000000	0.005847	0.021080	
Risk Profile	0.005847	1.000000	0.014037	
Previous Claims History	0.021080	0.014037	1.000000	
Credit Score	0.003211	-0.013158	-0.002146	



#### P-values Matrix:

	Customer ID	Age	Income Level	Location	\
Customer ID	NaN	0.001107	0.014758	0.004849	
Age	0.001107	NaN	0.425332	0.937776	
Income Level	0.014758	0.425332	NaN	0.001404	
Location	0.004849	0.937776	0.001404	NaN	
Claim History	0.000068	0.680304	0.003327	0.000433	
Coverage Amount	0.380881	0.118685	0.000031	0.000605	
Premium Amount	0.097115	0.396542	0.63945	0.87772	
Deductible	0.853501	0.000289	0.009759	0.003469	
Risk Profile	0.013993	0.0	0.003859	0.729175	
Previous Claims History	0.155216	0.176958	0.270748	0.461708	
Credit Score	0.000019	0.713173	0.000005	0.150104	

	Claim History	Coverage Amount	Premium Amount	\
Customer ID	0.000068	0.380881	0.097115	
Age	0.680304	0.118685	0.396542	
Income Level	0.003327	0.000031	0.63945	
Location	0.000433	0.000605	0.87772	
Claim History	NaN	0.93804	0.000004	
Coverage Amount	0.93804	NaN	0.703299	
Premium Amount	0.000004	0.703299	NaN	
Deductible	0.954256	0.075863	0.786979	
Risk Profile	0.128416	0.041563	0.000649	
Previous Claims History	0.000065	0.005277	0.0	
Credit Score	0.531375	0.913755	0.002651	

	Deductible	Risk Profile	Previous Claims History	\
Customer ID	0.853501	0.013993	0.155216	
Age	0.000289	0.0	0.176958	
Income Level	0.009759	0.003859	0.270748	
Location	0.003469	0.729175	0.461708	
Claim History	0.954256	0.128416	0.000065	
Coverage Amount	0.075863	0.041563	0.005277	
Premium Amount	0.786979	0.000649	0.0	
Deductible	NaN	0.176248	0.000001	
Risk Profile	0.176248	NaN	0.001166	
Previous Claims History	0.000001	0.001166	NaN	
Credit Score	0.457653	0.002338	0.619649	

Claim History & Premium Amount: A weak negative correlation (-0.0199) suggests that as claim history increases, the premium amount may slightly decrease.

Age & Risk Profile: A very weak negative correlation (-0.0239) suggests that older individuals may have a slightly lower risk profile.

### Conclusion:

Through this experiment, we gained valuable insights into D3.js and its powerful capabilities for data visualization. We explored how to effectively plot various types of graphs, including bar charts, line plots, histograms, and more.