FINAL PROJECT

GradVizz: Decoding Engineering Careers

Aneesh Sambu - 2023121012
Pavitra Pinninti - 2022111029
Paridhi Jain - 2022101119

Brief Description of Data :

Our project aims to explore various aspects of job market dynamics and educational trends through data visualization. By analyzing datasets containing information on job titles, regional hiring patterns, salary distributions, educational performance, and gender representation, we seek to uncover valuable insights that can inform decision-making processes in career planning, recruitment strategies, and educational policies. Through interactive visualizations, we delve into questions surrounding job title distributions, regional disparities in hiring trends, salary differentials across cities, correlations between academic achievements, and gender distributions within the dataset. These visualizations offer a comprehensive overview of the job market landscape and educational landscape, facilitating a deeper understanding of the factors influencing career trajectories and educational outcomes.Following are the parameters that our dataset contains:

- ID
- Salary
- Designation
- Job City
- Gender
- Degree
- Specialisation
- College GPA
- College State
- Graduation Year
- English
- Logical
- Quant
- Domain
- Computer Programming
- ElectronicsAndSemicon
- Computer Science
- Mechanical Engg
- Electrical Engg

- Telecom Engg
- Civil Engg
- Conscientiousness
- Agreeableness
- Extraversion
- Neuroticism
- Openness to experience
- 10th percentage
- 12th percentage

The dataset comprises approximately 40 independent variables and 4000 data points, with each candidate having a unique identifier.

The dataset used would be the AMCAT (AMEO) 2015 dataset from Kaggle: https://www.kaggle.com/code/pavanyeluri/amcat-employability-outcome-eda

Progress

Phase 1

During Phase 1, we selected our topic and dataset focusing on graduates after the AMCAT exam. We outlined the key questions we aimed to address through our visualizations:

1. Which personality traits (including extraversion, openness, conscientiousness, and agreeableness) significantly influenced employability?
2. What are the typical job titles and job locations for engineering graduates with AMCAT scores?
3. Are there variations in hiring patterns across different regions?
4. What is the distribution of salaries among engineering graduates with AMCAT scores?
5. 4. Which industries or domains are most popular among engineering graduates with AMCAT scores?

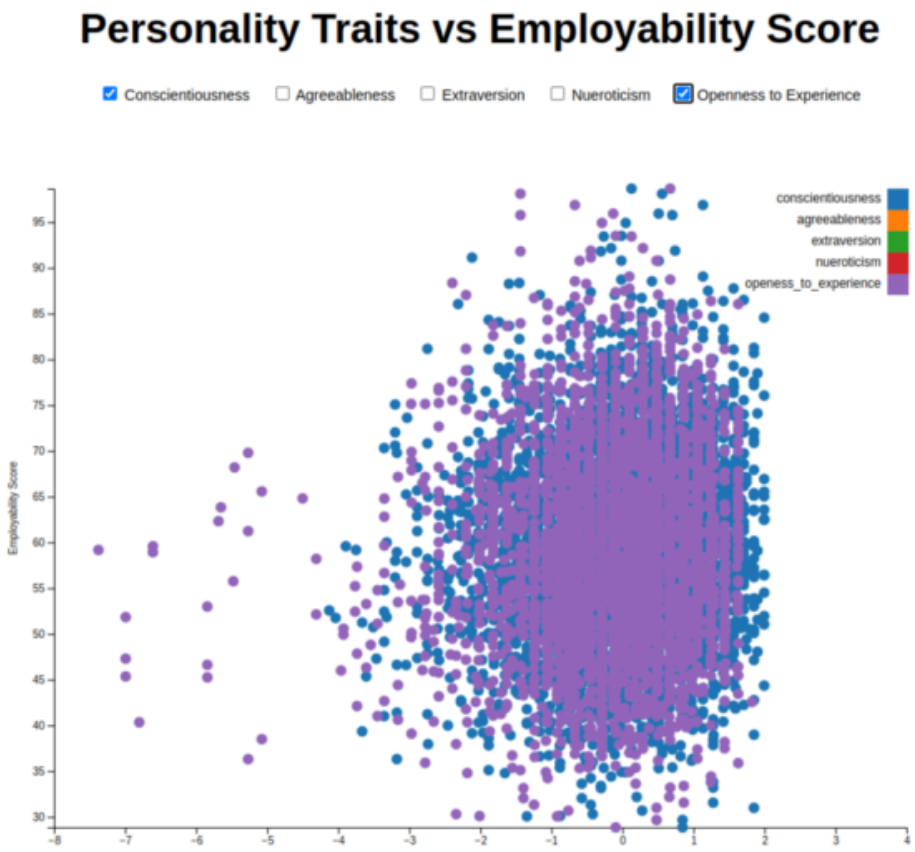
These questions formed the basis of our analysis and guided the development of our visualizations to uncover insights from the dataset.

Phase 2

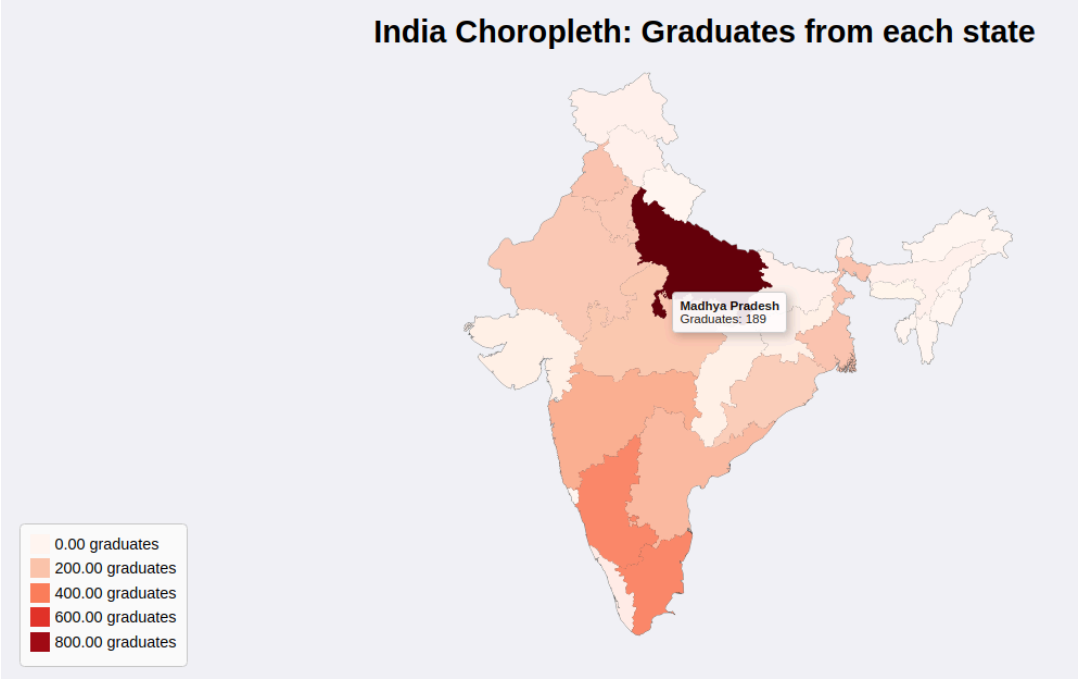
In Phase two, we opted to address two questions outlined in Phase 1.

Firstly, we aimed to explore the influence of personality traits, such as extraversion, openness, conscientiousness, and agreeableness, on employability. Initially, we planned to visualize this relationship using a scatterplot with radio buttons for toggling between different

personality traits. However, upon evaluation, we found this visualization unsuitable for in-depth analysis and subsequently abandoned it.



Secondly, we tackled the question of regional hiring patterns by creating a choropleth map. This map allows users to hover over each state to view the state's name and the number of graduates from that particular region.



Phase 3

Questions Addressed in Our Visualizations:

The questions we plan on visualizing are as follows:

1. What are the top 10 Job Titles and how are they distributed?
2. Do hirings vary from region to region?
3. Which cities offer the highest paying jobs?
4. Is there a correlation between college GPA and 12th percentage?
5. What is the distribution of average scores among the individuals?
6. What is the gender distribution among the individuals in the dataset?

Job Titles Distribution by Gender and Salary

The attributes utilized for this visualization include Designation, Gender, and Salary.

This visualization utilizes a combination of visual encoding techniques to effectively present insights into gender distribution and salary differentials across the top 10 job titles.

Visual Encoding:

- Bar chart: Represents the count of male and female individuals for each job title.
- Line plot: Overlays the average salary for each job title, with blue markers indicating male average salaries and orange markers indicating female average salaries.
- Color: Blue and orange colors are used to differentiate between male and female average salaries, respectively, enhancing visual distinction and clarity.

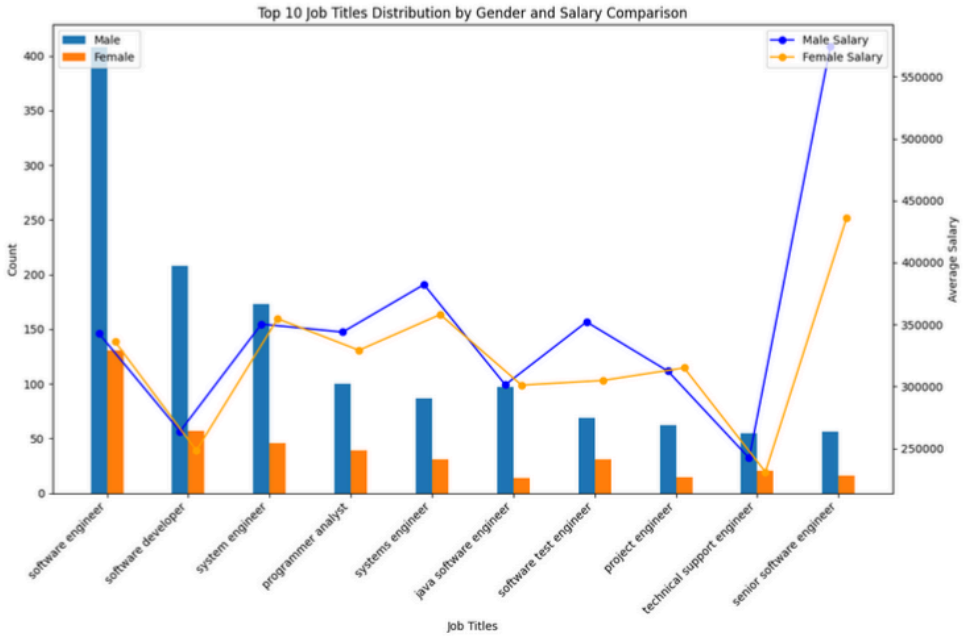
Task and Users:

- Specific tasks that can be accomplished with this visualization include:
    1. Identifying gender distribution within specific job titles.
    2. Comparing average salaries between male and female individuals across different job roles.
    3. Discerning any disparities in job title representation between genders.
- Users of this visualization may include policymakers, HR professionals, researchers, and advocates for gender equity in the workforce. They can use the insights gained from this visualization to inform discussions, initiatives, and policies aimed at promoting gender equality and addressing disparities in pay scales across various job roles.

Limitations:

- While this visualization effectively presents gender distribution and salary differentials, it may not provide insights into the underlying factors contributing to these disparities, such as systemic biases or occupational segregation.

- Additionally, the visualization's reliance on average salaries may obscure individual-level variations in earnings, potentially oversimplifying the complexity of gender-based salary discrepancies in the workforce.



Regional Hiring Trends

The attribute utilized for this visualization is College State.

This visualization utilizes a choropleth map to visually represent hiring patterns across different regions, with a focus on the number of graduates in each state. Key elements of the visualization include:

Visual Encoding:

- Choropleth map: The map employs color gradients to depict variations in the number of graduates across states, with darker shades indicating higher numbers and lighter shades indicating lower numbers.
- Color: The color gradient effectively communicates the distribution of graduates, allowing viewers to easily identify states with high or low numbers of graduates.

Interactivity:

- Hovering capabilities: Users can hover over each state to reveal specific information, such as the state name and the corresponding number of graduates. This interactive feature enhances user engagement by providing direct access to detailed information about each state.

Task and Users:

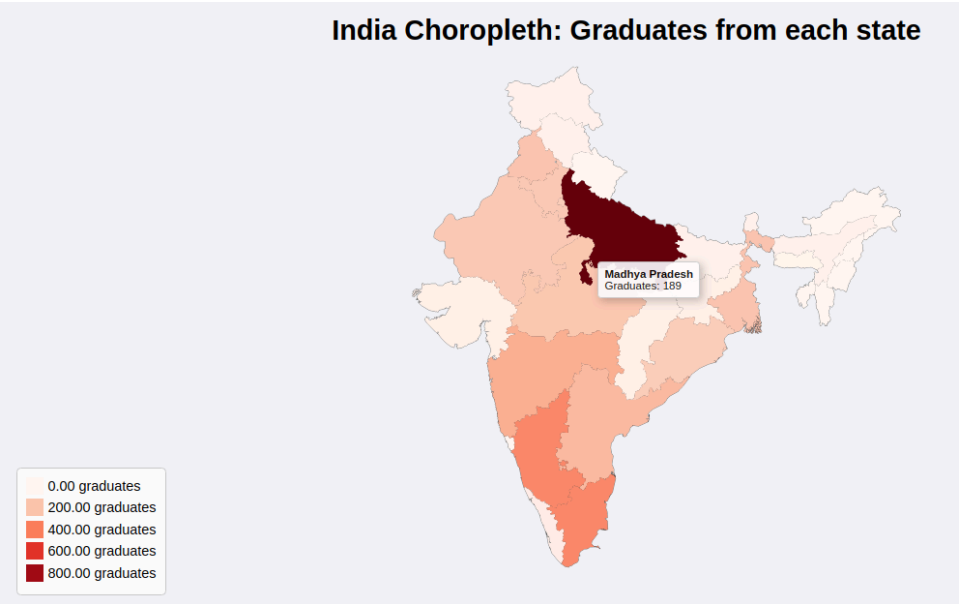
- Specific tasks that can be accomplished with this visualization include:
    1. Identifying regional disparities in hiring patterns based on the number of graduates.
    2. Observing geographical trends in employment opportunities across states.
    3. Exploring detailed information about each state's graduate population by hovering over individual states.
- Users of this visualization may include policymakers, recruiters, and researchers interested in understanding and addressing regional variations in hiring practices. They can use the insights gained from this visualization to inform decision-making processes related to workforce planning, recruitment strategies, and educational policies.

Effectiveness:

- By employing a choropleth map with hovering capabilities, this visualization effectively communicates complex information about regional hiring patterns in a visually intuitive manner.
- The color gradient and interactivity enhance user engagement and facilitate the exploration of detailed information about each state's graduate population.

Limitations:

- While the visualization provides valuable insights into regional hiring patterns, it may not capture other factors influencing employment opportunities, such as industry sectors or job market dynamics.
- Users should interpret the visualization with caution and consider additional contextual information when making decisions or drawing conclusions about regional variations in hiring practices.



Correlation between College GPA and 12th Percentage

The attributes utilized for this visualization include College GPA and 12th percentage

The scatter plot utilizes visual encoding, navigation, and color to effectively illustrate the relationship between 12th percentage and college GPA:

Visual Encoding:

- Scatter plot: Each data point represents an individual, with their 12th percentage plotted on the y-axis and college GPA on the x-axis. This visual encoding allows viewers to easily identify the relationship between the two variables.

Navigation:

- Radio buttons: Users can switch between different color schemes, enabling them to customize the visualization according to their preferences. This navigation feature enhances user interaction and facilitates exploration of data patterns.

Color:

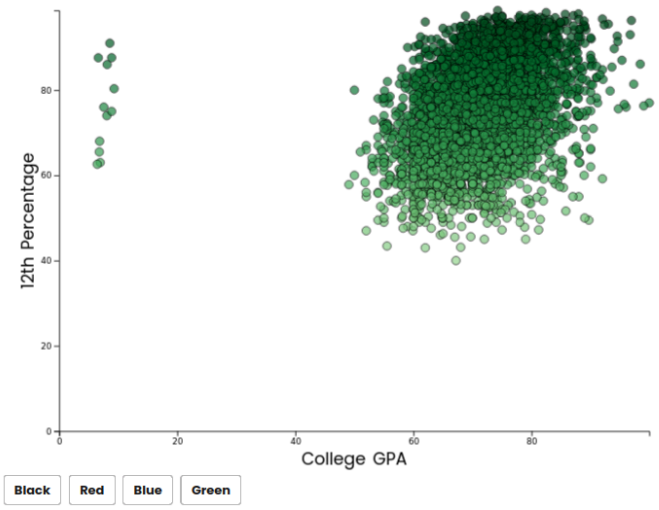
- Color variation: Different color schemes are employed to represent data points, providing users with flexibility in visualizing the data. For example, users can toggle between colors representing different groups or categories, such as gender or age. The use of color enhances visual appeal and aids in distinguishing between data points.

Tasks and Users:

- Specific tasks that can be accomplished with this visualization include:
    1. Identifying the correlation between 12th percentage and college GPA.
    2. Exploring how different color representations affect the visualization of data patterns.
- Users of this visualization may include researchers, educators, and policymakers interested in understanding the relationship between academic performance in high school and college. They can use this visualization to gain insights into factors influencing college GPA and to identify potential areas for intervention or support.

Limitations:

- While the scatter plot effectively illustrates the correlation between 12th percentage and college GPA, it may not capture other factors influencing academic performance, such as socioeconomic status or extracurricular activities.
- Users should interpret the visualization with caution and consider additional contextual information when drawing conclusions about academic achievement.



Top 10 Highest Paying Job Cities

The attributes utilized for this visualization include Salary and Job Cities

The pie chart visualization employs visual encoding, navigation, and color choices to effectively convey information about salary distributions among the top 10 highest paying job cities. Key elements of the visualization include:

Visual Encoding:

- Pie chart: The primary visual element used to represent salary distributions across different cities. Each segment of the pie chart corresponds to a specific city, with the size of the segment indicating the proportion of average salary attributed to that city.
- Color gradient: Different colors are used to distinguish between segments, making it easier for viewers to differentiate between cities and identify patterns in salary distributions.

Navigation:

- On-hover details: Upon hovering over each segment, viewers can see the city name and average salary of individuals in that city, providing additional context without cluttering the main display.

Color:

- Color gradients are chosen to enhance visual clarity and make it easier for viewers to distinguish between different segments of the pie chart.
- The use of contrasting colors helps draw attention to disparities in salary levels between cities, with larger segments representing cities with higher average salaries and smaller segments indicating cities with lower average salaries.

Specific Tasks:

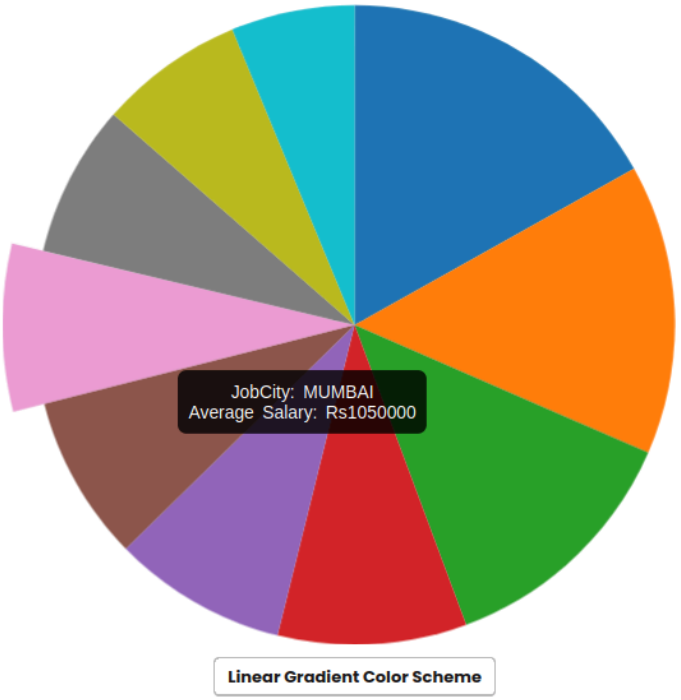
- Identify cities with the highest average salaries: Viewers can easily identify cities with the largest segments in the pie chart, indicating higher average salaries.
- Compare salary levels across cities: The varying sizes of the segments allow viewers to compare salary levels among different cities and identify potential disparities.
- Gain insights into salary trends and disparities: By visually inspecting the pie chart, viewers can gain valuable insights into salary distributions across key job locations and make informed decisions in their career pursuits.

Users:

- Users of this visualization may include job seekers, recruiters, policymakers, and researchers interested in understanding salary trends and disparities across different urban centers.
- Recruiters and employers can use the visualization to inform hiring strategies and compensation packages based on regional salary trends.
- Job seekers can use the visualization to identify cities where professionals typically command higher salaries and tailor their job search accordingly.

Limitations:

- While the pie chart visualization offers a straightforward and visually engaging representation of salary distributions, it may oversimplify complex salary dynamics and fail to capture nuances such as industry-specific salary variations or cost-of-living considerations.



Gender Distribution

The attribute utilized for this visualization is Gender

Visual Encoding:

- Bar chart: The length of each bar corresponds to the number of individuals in each gender category, providing a clear visual representation of gender distribution.
- Hover functionality: Users can hover over each bar to reveal a tooltip displaying the precise count of individuals for that gender category, enhancing interactivity and providing detailed information.

Color:

- Distinct colors may be used for the bars representing different genders, such as dark blue for males and light blue for females, to aid in differentiation and enhance visual clarity.

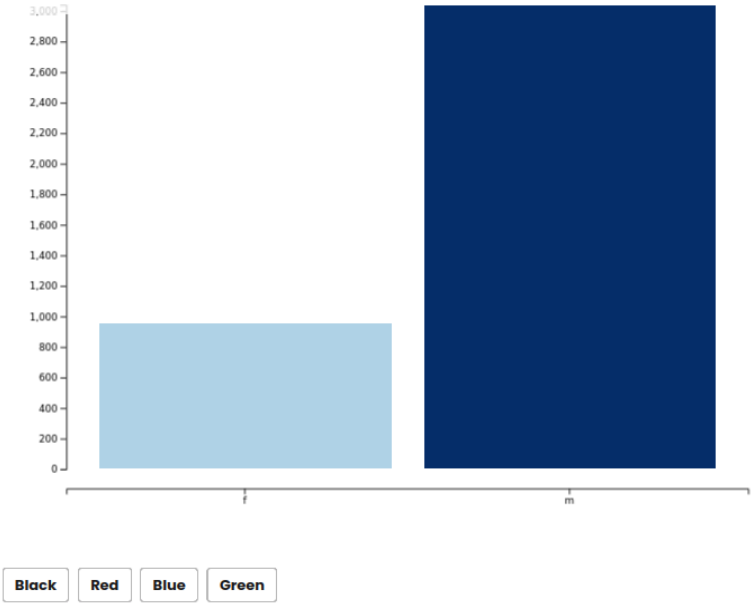
Task and Users:

Task: The visualization enables users to understand the gender distribution among individuals in the dataset and identify any disparities or patterns.

Users: Potential users of this visualization include researchers, policymakers, and analysts interested in exploring gender demographics within a dataset. Recruiters or HR professionals may also use this visualization to analyze workforce diversity.

Limitation:

While the bar chart effectively communicates gender distribution, it may not capture other factors influencing demographics, such as age, ethnicity, or socioeconomic status. Users should interpret the visualization in conjunction with other demographic data to gain a comprehensive understanding of the dataset's composition.



Average Score Distribution

The attribute utilized for this visualization is AverageScore

Visual Encoding:

- Histogram: The histogram visualizes the distribution of average scores, with bars representing the frequency of scores falling within each range or bin. The x-axis denotes "Average Score," and the y-axis indicates "Frequency."

Navigation:

- Hovering Functionality: Users can hover over individual bars to reveal specific information, such as the frequency of scores within each range or bin.

Color:

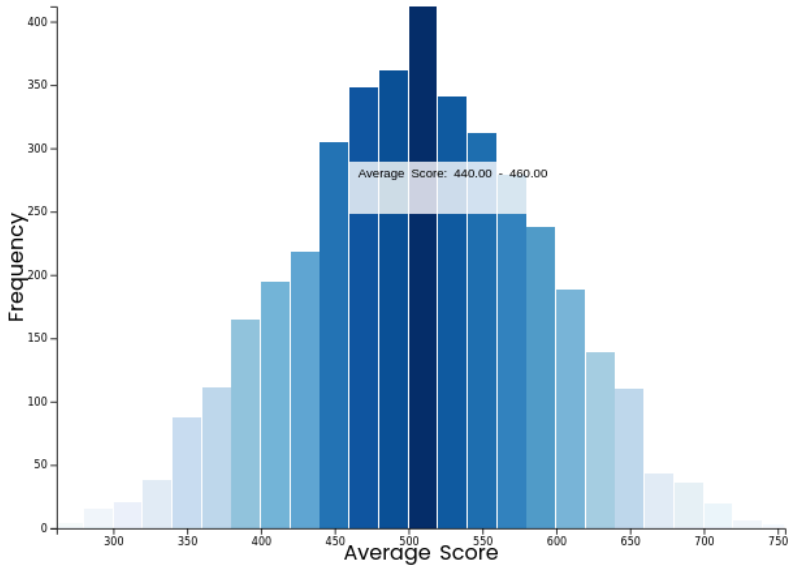
- Colors may be used to differentiate the bars or highlight specific aspects of the distribution, such as shading bars to indicate higher or lower frequency ranges. A single color scheme may suffice to maintain clarity.

Task and Users:

- Task: The visualization enables users to understand the distribution of average scores within the dataset, including insights into the shape, range, and spread of scores.
- Users: Researchers, educators, and policymakers interested in assessing the performance patterns of individuals in the dataset may find this visualization valuable. Educational institutions or testing organizations may also utilize it to evaluate test performance and identify areas for improvement.

Limitation:

- While the histogram provides valuable insights into the distribution of average scores, it may not capture underlying factors influencing performance, such as test difficulty, study habits, or socioeconomic background. Users should interpret the visualization in conjunction with additional contextual information to gain a comprehensive understanding of performance patterns.



Black Red Blue Green

Gradient Magnitude: X-Axis

Link for the video:

https://youtu.be/ALhJJS7Q1jU

Contributions:

All decided the visualizations

Aneesh Sambu:

- Visualisations
- Website
- Video

Paridhi Jain:

- Visualization
- Documentation
- video

Pavitra Pinninti:

- Visualization
- Data Cleaning
- Video