[Document title]

[Document subtitle]

Contents

[Data Import and Exploration: 2](#_Toc135428088)

[R Markdown Analysis (if applicable): 2](#_Toc135428089)

[Integration of Power BI and R (if applicable): 2](#_Toc135428090)

[Interpretations and Documentation: 3](#_Toc135428091)

[Data Transformation 3](#_Toc135428092)

[Visualization Interpretations 4](#_Toc135428093)

[Pie Chart: 4](#_Toc135428094)

[Slicer: 4](#_Toc135428095)

[Funnel Chart: 5](#_Toc135428096)

[Gauge: 5](#_Toc135428097)

[Scatter chart: 6](#_Toc135428098)

The methodologies involved in combining Power BI and R Markdown, or integrating R visualizations in Power BI, follows these steps:

# Data Import and Exploration:

* We identified the data sources and import them into Power BI or R.
* We performed exploratory data analysis (EDA) to understand the data structure, identify patterns, and detect outliers.
* Cleanse and preprocess the data as needed.
* Power BI Analysis (if applicable):

Building interactive visualizations, reports, and dashboards using Power BI's interface.

We utilized Power BI's native capabilities for data modeling, filtering, and creating calculated fields.

Apply various visualizations, such as charts, tables, maps, and custom visuals, to represent the data.

We use Power BI's interactive features like drill-through, slicers, and filters to explore the data and derive insights.

# R Markdown Analysis (if applicable):

We already have installed R and RStudio on our machines if not already done.

Loading the required libraries/packages for our analysis.

Write R code within an R Markdown file (.Rmd) to perform specific analysis tasks.

Import the data from the appropriate source into R.

Apply statistical analysis techniques, data transformations, or custom visualizations using R code.

# Integration of Power BI and R (if applicable):

We could also follow the second approach, integrating R visualizations within Power BI:

Using the "R Visual" component in Power BI to embed R scripts and generate visualizations.

Write R code directly within Power BI to create custom visuals or perform advanced analysis.

Combine Power BI's native visuals with R visuals to present a comprehensive analysis.

# Interpretations and Documentation:

By analyzing the results obtained from both Power BI and R Markdown/R visuals we can derive insights, identify trends, patterns, correlations, or any other findings from the data.

Explain the interpretations, conclusions, and actionable recommendations based on the analysis.

# Data Transformation

The data transformation that had taken place in this project wasn’t that much firstly because the data and datasets that we use were already cleaned and trimmed thus the only logical steps for us was to find data in the sets/ dataset to use.

We needed to find the correct and appropriate data for us to use in our analysis and thus some data was added, and some was removed to make it more convenient for the visualizations to be used and making it more understandable for anyone that looks at the report. It also makes it convenient for calculative columns to be created.

Transformation the took place was the Change on one of our dataset files from XML to CSV by just saving the file in the csv format or by using MikTex to knit the file into another language. MiKTex is an application that is also used while working in RStudio to change the file, language and format using the knit function.

We probably performed basic data transformations on a routine basis. When we convert a Microsoft Word file to a PDF, for example, we are transforming data. But data transformation plays a more important and sophisticated role in big data analytics. That is because when we are dealing with large volumes of data, different types of data analytics tools and different data storage systems, we are likely to encounter situations where a large amount of data needs to be transformed from one format to another.

According to my knowledge there we no other Unstructured to Structured Data Transformation done in the project because the data was already cleaned and transformed ready to use so thus no unstructured data was but there was a relationship that was created between the tables.

# Visualization Interpretations

# Pie Chart:

1. Clear Comparison: A pie chart is an effective way to visually represent proportions or percentages. By using a pie chart, you can clearly showcase the relative sizes of two categories - the number of students who attend nursery and those who don't. The chart will provide an immediate visual comparison between the two categories.
2. Simple and Intuitive: Pie charts are simple and intuitive to interpret. The chart divides a circle into segments, with each segment representing a category. The size of each segment directly corresponds to the proportion or percentage it represents. Can easily understand the comparison between the number of students attending nursery and those who do not at a glance.
3. Limited Data Points: When comparing only two categories, a pie chart is suitable because it effectively presents the data without cluttering the visual. Pie charts are particularly useful for displaying data with few categories, making it ideal for your comparison between attending nursery and not attending nursery.

# Slicer:

1. Easy Data Exploration: By using a slicer, we enable easy exploration of the data across different race/ethnicity groups. Users can simply select the desired group from the slicer, and the visualizations will dynamically update to show the corresponding data. This interactivity enhances the user experience and allows for quick comparisons and analysis.
2. Flexibility and Customization: Slicers in Power BI offer flexibility and customization options. We can choose from various slicer types, such as a dropdown, list, or checkbox, based on your preference and requirements. We can also customize the appearance and behavior of the slicer to align with our visual design and user interaction preferences.
3. Clear Data Filtering: It provides a clear and intuitive way to filter data based on race/ethnicity groups. Users can easily see the available options and make selections, without the need for complex filtering mechanisms. This makes it simpler for users to focus on specific groups of interest and analyze the corresponding data.
4. Enhanced Data Insights: By using a slicer to visualize race/ethnicity groups, you enable better insights into the distribution and patterns within each group. Users can explore how different variables or metrics vary across groups, identifying any disparities or trends that may exist. This can be particularly useful when analyzing diversity, equity, and inclusion efforts or studying demographic data.

# Funnel Chart:

1. Clear representation: The funnel chart provides a visually intuitive representation of data, allowing us to easily understand the distribution of absence students across different free time categories. The decreasing width of the funnel segments effectively showcases the gradual reduction in student count as you move from one free time category to another.
2. Highlighting drop-off points: The Funnel charts help identify significant drop-off points in the absence student count. By analyzing the narrowing funnel segments, we can pinpoint the free time categories where the largest number of students are absent and where the drop-offs occur. This information can be crucial for identifying potential issues or areas of concern that may require further investigation or intervention.
3. Comparative analysis: The varying widths of the funnel segments allow for easy comparison between different free time categories. You can quickly identify which free time categories have higher or lower absence rates by comparing the width of the segments. This comparative analysis enables us to understand the relative impact of different free time categories on student absences.
4. Focus on key metrics: Funnel charts are particularly useful for highlighting key metrics and trends. By focusing on the funnel's top and bottom sections, we can easily identify the free time categories with the highest and lowest absence rates. This information can help you prioritize interventions or allocate resources effectively based on the most critical areas.

# Gauge:

1. Clear and concise representation: The Gauge visualization offers a simple and intuitive way to present the overall number of failures. It uses a visual indicator, such as a dial or a progress bar, to convey the information effectively.
2. Instant data interpretation: The visual nature of the Gauge allows users to quickly grasp the magnitude of failures without needing to analyze raw data or read lengthy reports. It provides an immediate snapshot of the situation.
3. Focus on key metrics: The Gauge visualization can be customized to highlight specific thresholds or targets related to failure rates. This helps stakeholders and decision-makers easily identify whether the number of failures is within acceptable limits or if action needs to be taken.
4. Comparative analysis: By incorporating multiple Gauges or using dual-axis Gauges, you can compare the current failure count with historical data or set benchmarks. This enables trend analysis and helps identify patterns or anomalies in failure rates over time.
5. Visualizing progress towards goals: Gauges can be configured to display progress towards predefined targets or goals. This provides a visual indication of how close the organization is to achieving its objectives related to reducing failures.

# Scatter chart:

1. Identifying Correlations: By using a scatter chart, you can easily identify correlations between the math score and writing score. Correlations can be positive (both scores increase together), negative (one score increases while the other decreases), or no correlation (no clear relationship between the scores). Analyzing the scatter plot can help one identify the strength and direction of the correlation.
2. Outlier Detection: Scatter charts are effective in detecting outliers, which are data points that deviate significantly from the general pattern. Outliers can provide valuable insights into the data and potentially indicate areas for further investigation. By plotting the math score against the writing score, you can quickly identify any extreme values that fall outside the usual range.
3. Segmentation and Clustering: Scatter charts can assist in segmenting or clustering the data based on specific criteria. For example, one may want to analyze how different groups of students perform based on their math and writing scores. By assigning different colors or shapes to the data points representing different student groups, one can then easily observe any patterns or differences between the segments.
4. Trend Analysis: Scatter charts can help in identifying trends over time or any changes in the relationship between the math and writing scores. By adding a trendline to the scatter chart, we can visualize the overall direction of the relationship. This can be useful for making predictions or forecasting future outcomes based on historical data.
5. Interactive Data Exploration: Power BI allows for interactivity in scatter charts. Users can select data points, filter the data based on specific criteria, zoom in or out, and explore the underlying data further. This interactivity enhances data exploration and facilitates deeper analysis of the relationship between the math and writing scores.
6. Integration with Other Visualizations: Scatter charts can be combined with other visualizations within Power BI to provide a comprehensive view of the data. For example, we can include filters, slicers, or drill-through functionality to interact with other charts or visuals in