Unlocking Insights: Data Visualization with Tableau

Author: Quy Nguyen Cong

Date: 22nd March 2024

Introduction

Visualization plays a pivotal role in analyzing the results of data-driven projects, offering insights that may not be immediately apparent from raw data alone. By visually representing complex datasets, visualization tools like Tableau allow us to explore trends, patterns, and relationships, facilitating a deeper understanding of the underlying data. This visual interpretation enhances decision-making processes, enabling stakeholders to identify key insights and formulate actionable strategies based on data-driven evidence.

The Power of Tableau

Tableau is a powerful and versatile data visualization tool that empowers users to create interactive and engaging visualizations from various datasets. Its user-friendly interface, coupled with advanced features for data exploration and analysis, makes it an indispensable tool for professionals across industries. With Tableau, users can effortlessly create insightful dashboards, charts, and graphs, enabling them to communicate complex data concepts effectively and drive informed decision-making.

The Importance of Probability Prediction

In this analysis, the use of probability prediction offers distinct advantages over binary categorization (e.g., category 1 or 0). While binary categorization provides a straightforward classification of outcomes, probability prediction offers a more nuanced understanding of the likelihood of specific events occurring. By predicting probabilities, we gain insights into the relative likelihood of different outcomes, allowing for more granular analysis and decision-making. This probabilistic approach enables stakeholders to assess risks, prioritize actions, and optimize resource allocation based on the likelihood of specific outcomes, ultimately leading to more informed and effective strategies. Therefore, incorporating probability prediction into our analysis provides a richer and more comprehensive understanding of the underlying data and its implications.

Building on the importance of visualization and probability prediction outlined in the introduction, let's delve into three key analyses to gain deeper insights into absenteeism patterns.

Insight 1: Relationship between Age and Probability of Absence

The analysis reveals a generally positive correlation between age and the predicted probability of absenteeism. As employees age, the model predicts a higher likelihood of absenteeism. For instance, an individual aged 30 has only a 28% chance of being absent from work for more than three hours, whereas someone aged 50 has nearly an 80% chance of excessive absenteeism. However, it's important to note that age is not the sole determining factor, as indicated by the spread of data points across the age spectrum.

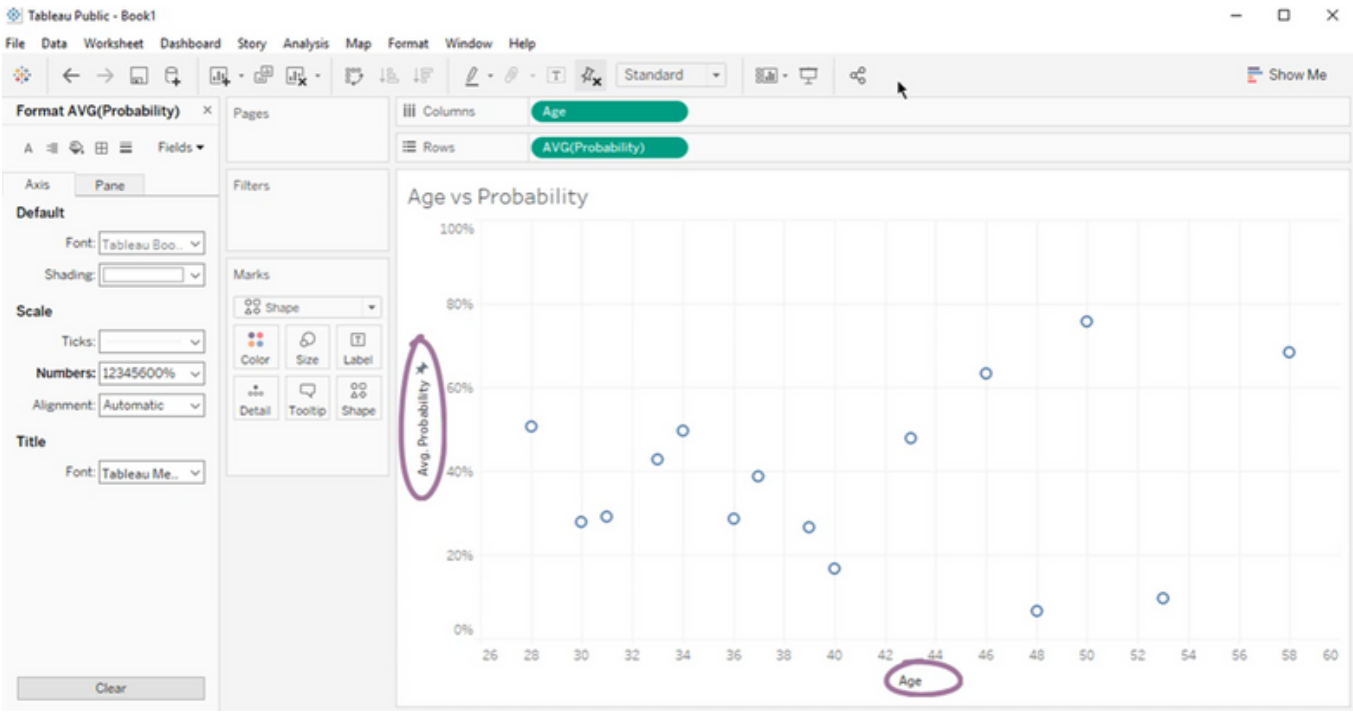

Fig 1. Relationship between Age and Probability of Absence

Insight 2: Reason for Absence and Probability of Absence

Group 1 reasons, representing serious illnesses, are associated with more prolonged and significant disruptions to health, leading to a higher likelihood of exceeding the 3-hour absenteeism threshold. Conversely, Group 4 reasons, characterized by lighter issues such as dental appointments or medical consultations, are less likely to result in extended absences due to their lower severity and shorter appointment durations. The analysis also highlights the impact of sample size, with limited observations for Reason 3 making it challenging to draw definitive conclusions about its effect on absence duration.

In conclusion, the analysis underscores a correlation between the severity of the reason for absence and the probability of exceeding the 3-hour threshold. Serious illnesses (Group 1) are more likely to cause extended absences, while minor reasons (Group 4) are less disruptive and have a lower probability of exceeding the limit.

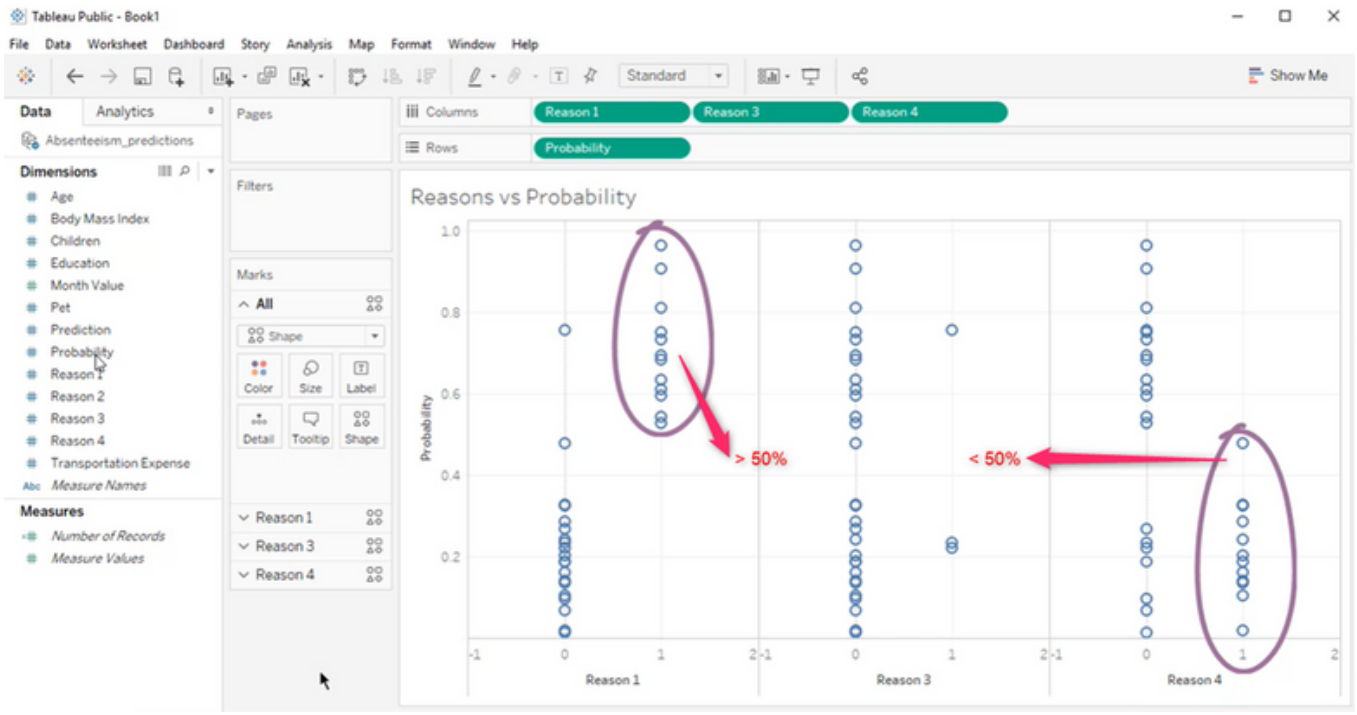

Fig 2. Relationship between Reason for Absence and Probability

Insight 3: Transportation Expense, Number of Children, and Probability of Absence

The Tableau worksheet visualizes data on transportation expense, number of children, and the probability of excessive absenteeism. A scatter plot illustrates a weak positive correlation between transportation expense and the probability of excessive absence, suggesting that employees with higher transportation expenses may be slightly more likely to be excessively absent. However, the correlation is not strong, and the data is scattered.

Additionally, the size of circles in the scatter plot represents the number of children an employee has. Despite variations in circle sizes, there doesn't appear to be a significant impact of the number of children on the probability of excessive absenteeism.

Overall, the visualization suggests that transportation expense may serve as a weak indicator of absenteeism, while the number of children does not significantly influence the probability of excessive absenteeism.

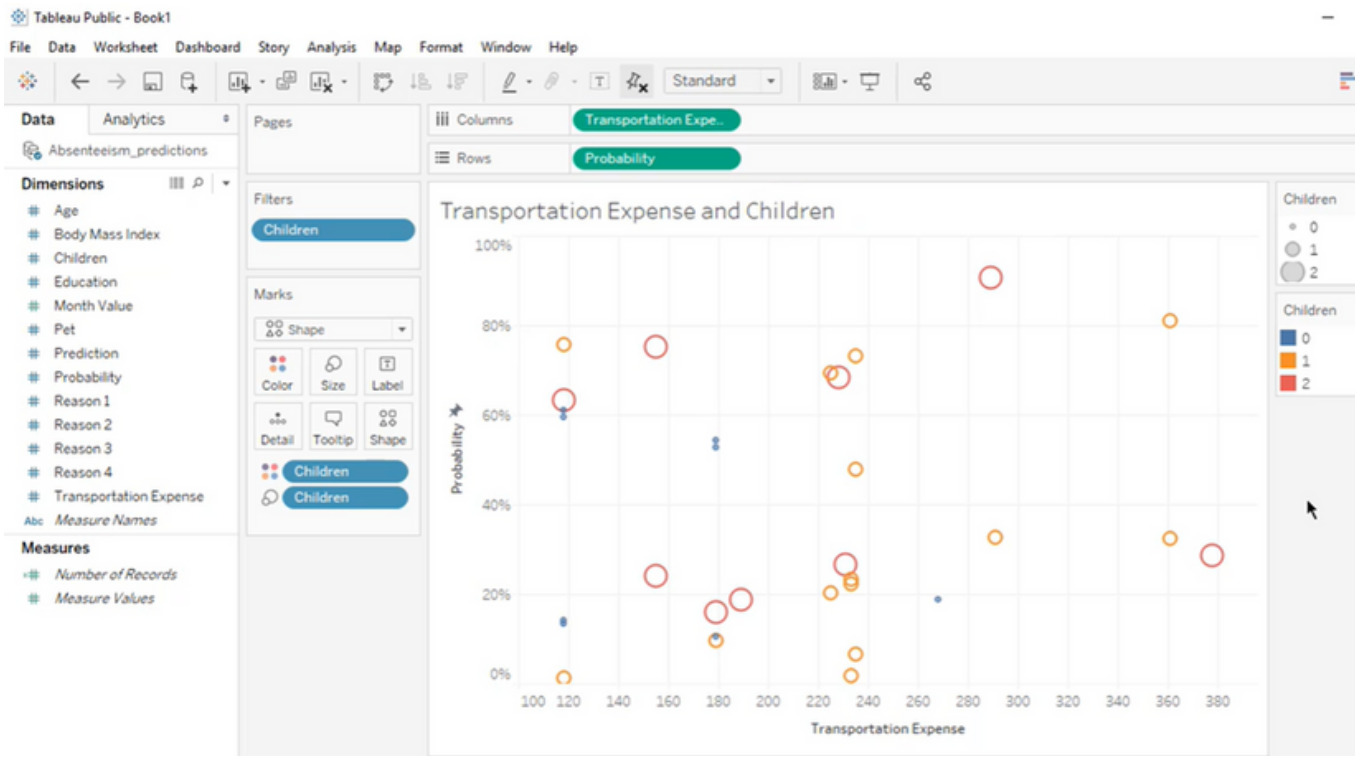

Fig 3. Relationship between Transportation Expense, Number of Children, and Probability of Absence

In summary, our analyses provide valuable insights into absenteeism patterns, highlighting correlations between various factors such as age, reasons for absence, and demographic variables. Leveraging visualization and probability prediction techniques, we've equipped stakeholders with actionable insights to address absenteeism effectively and foster a more productive work environment.

Thank for reading