1. Project overview:
    a. By analyzing customer behaviors, they seek to identify high-risk and low-risk groups more accurately. This helps in pricing policies, setting appropriate premiums, and managing claims.
    b. Gaining a competitive edge relies on data-driven decision-making. Insights into customer behaviors can lead to better customer retention, reduced claim costs, and ultimately, improved profitability.
    c. This project is of great significance to the insurance company as it directly impacts their ability to stay competitive and adapt to changing market dynamics. The insights derived from the data analysis will guide their future strategies and operational decisions.
2. Data Introduction:
    a. The dataset used in this project is a blend of real-world data and generated data specifically curated for analysis.
    b. In addition to the real-world data, a portion of the dataset has been deliberately generated to augment the available information. This generated data is designed to simulate various scenarios, behaviors, and risk profiles that may not be fully represented in the actual historical records.
3. Visual 1
    a. This suggests a strong correlation between speeding violations and a reduced likelihood of insurance claims.
    b. For the car insurance company, this finding has significant implications. It implies that the insurance company might consider adjusting their pricing strategies or risk assessments for customers with a history of speeding violations.
4. Visual 2:
    a. For the car insurance company, this finding holds important implications for risk assessment and pricing strategies.
    b. They may consider offering lower premiums or more favorable terms to customers with newer vehicles, reflecting the reduced claim probability associated with this group.
5. Strengths:
    a. Insurance datasets often suffer from class imbalance, with fewer claimants compared to non-claimants.
    b. Our model has been specifically designed to handle such imbalanced data effectively. It avoids biases that could result from skewed class distribution and ensures accurate predictions for both claim and non-claim cases.

6. Limitations:
    a. If there are inaccuracies or inconsistencies in the dataset, it could lead to suboptimal model outcomes. This is particularly relevant for the generated data, as the realism and accuracy of this data may impact the model's ability to generalize to real-world scenarios.
    b. Some customer behaviors, especially those driven by individual circumstances or unique life events, may not be fully captured by the model's features and algorithms.
7. Classification Consideration:
    a. **False Positives:** Resources are allocated for processing and investigating claims that turn out to be unnecessary, impacting efficiency and profitability.
    b. **False Negatives:** Genuine claims go unaddressed, leading to customer frustration and potential loss of trust in the insurance company.
8. Class Balance:
    a. **Accuracy:** In our context, it assesses how well we correctly identify genuine insurance claims among all predicted claims.
9. Final Recommendations:
10. Thank you:
    a. thank you for your attention and for listening to my presentation