**Title:**

**Predictive Modeling to Expand Vehicle Insurance Customer Base**

**Summary**

Our client is an insurance company that currently offers Health Insurance to its customers. They are now looking to expand their services and provide Vehicle Insurance as well. To make this expansion successful, they need our help in building a predictive model.

Insurance policies involve customers paying regular premiums to the company in exchange for coverage in case of specified losses, damages, illnesses, or accidents. For instance, you might pay an annual premium of Rs. 5000 for a health insurance plan that covers up to Rs. 200,000 in medical expenses. The question is, how can the insurance company afford to cover such high costs when they only charge a fraction of it as a premium? The answer lies in probabilities.

Not every policyholder will require a claim in a given year. While 100 customers may pay premiums, only a few of them (say 2-3) might actually need hospitalization. This way, the risk is spread across all policyholders, and everyone contributes to covering each other's potential losses.

Similarly, with Vehicle Insurance, customers pay premiums to receive compensation (called 'sum assured') in case of accidents or damage to their vehicles.

Building a predictive model to determine whether a customer is likely to be interested in Vehicle Insurance is crucial for the company's success. It allows the company to tailor its communication and outreach strategies to target the right customers, optimize its business operations, and increase revenue.

By analyzing this data, we can create a model that helps the company identify potential customers who would benefit from Vehicle Insurance, enabling them to make informed decisions and grow their business.

**Business Problem/Opportunity**

**Problem Statement**

Our client, an insurance company, seeks to broaden its product portfolio by introducing vehicle insurance to its existing health insurance policyholders. The primary challenge is to identify which policyholders are likely to be interested in purchasing vehicle insurance. This necessitates the development of a predictive model to optimize communication strategies, thereby increasing cross-selling opportunities and revenue.

**Key Opportunity**

The key opportunity in this endeavor lies in leveraging data-driven insights to identify and target potential customers for vehicle insurance. By tailoring marketing and communication strategies to these individuals, our client can maximize conversion rates and revenue growth while providing valuable insurance services.

**Specific Business Objectives**

1. **Identifying Potential Customers**: Develop a predictive model to identify health insurance policyholders likely to show interest in purchasing vehicle insurance.

2. **Effective Communication**: Tailor marketing and communication efforts to maximize conversion rates among potential customers.

**Process for Selecting and Gathering Data**

**Data Acquisition**

We obtained the dataset from Kaggle, which comprises customer information, including demographics, vehicle-related data, and policy-related details. Data gathering involved verifying the dataset's relevance to the business problem and ensuring it contained the necessary attributes for model development.

The raw datasets from Kaggle was comprised of train, test, and sample file so because of too much observations we only select and download the train file

The raw dataset from Kaggle contains approximately more than 3 lacs observation but when we apply data mining techniques data shrinks to 93 thousand observations

**Preliminary Data Exploration and Findings**

During the preliminary data exploration phase, we conducted the following activities:

* Analyzed the dataset's structure and size.
* Checked for missing values and addressed data quality issues.
* Explored the distribution of key variables.

These efforts provided valuable insights into the dataset, including demographic characteristics and initial observations regarding customers potentially interested in vehicle insurance.

**Data Preparation**

Data preparation involved several critical steps to ensure the dataset's suitability for modeling:

* **Balancing Technique**: We addressed class imbalance by applying resampling methods to prevent model bias.
* **Categorical to Numerical Conversion**: Categorical variables were transformed into numerical format using encoding techniques suitable for modeling.
* **Standardization**: Numerical features were scaled to ensure consistent model performance.

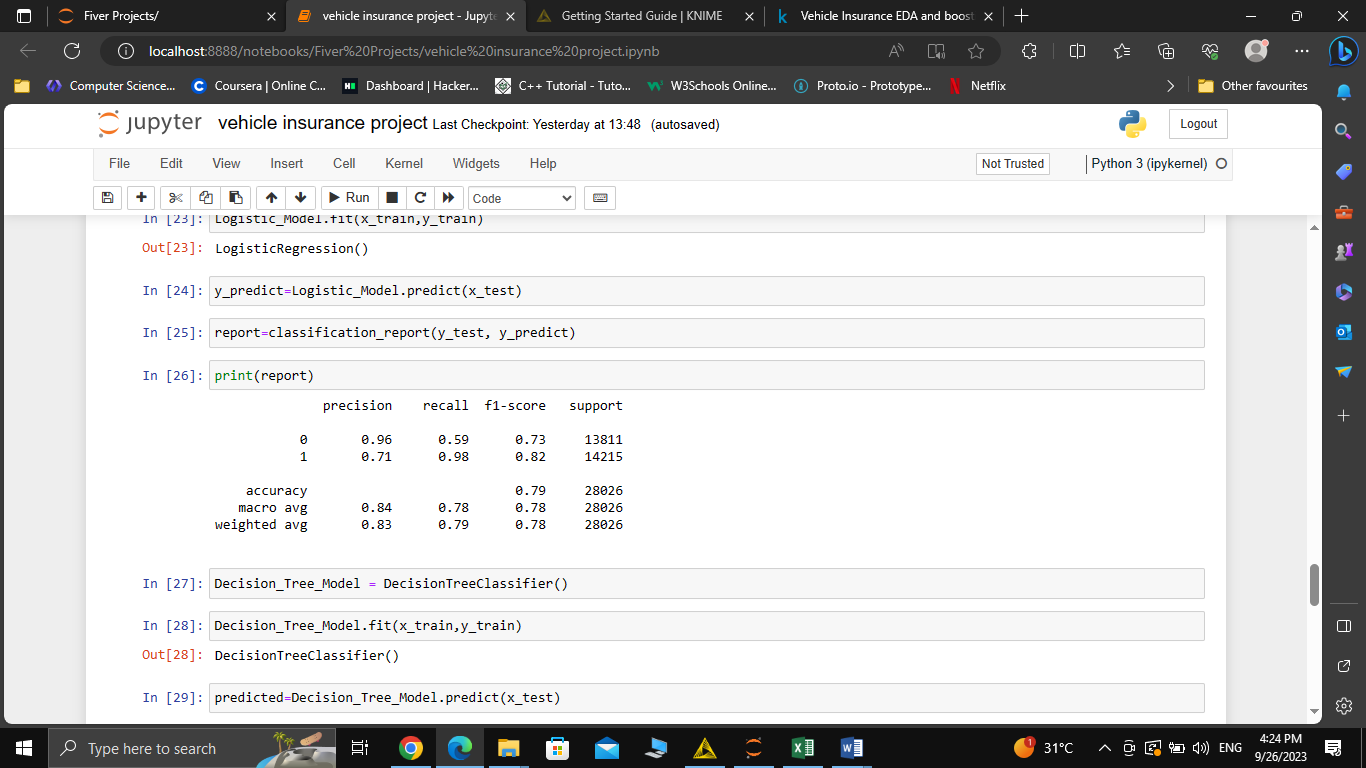
**Data Modeling and Mining Assessments**

**Model Selection and Training**

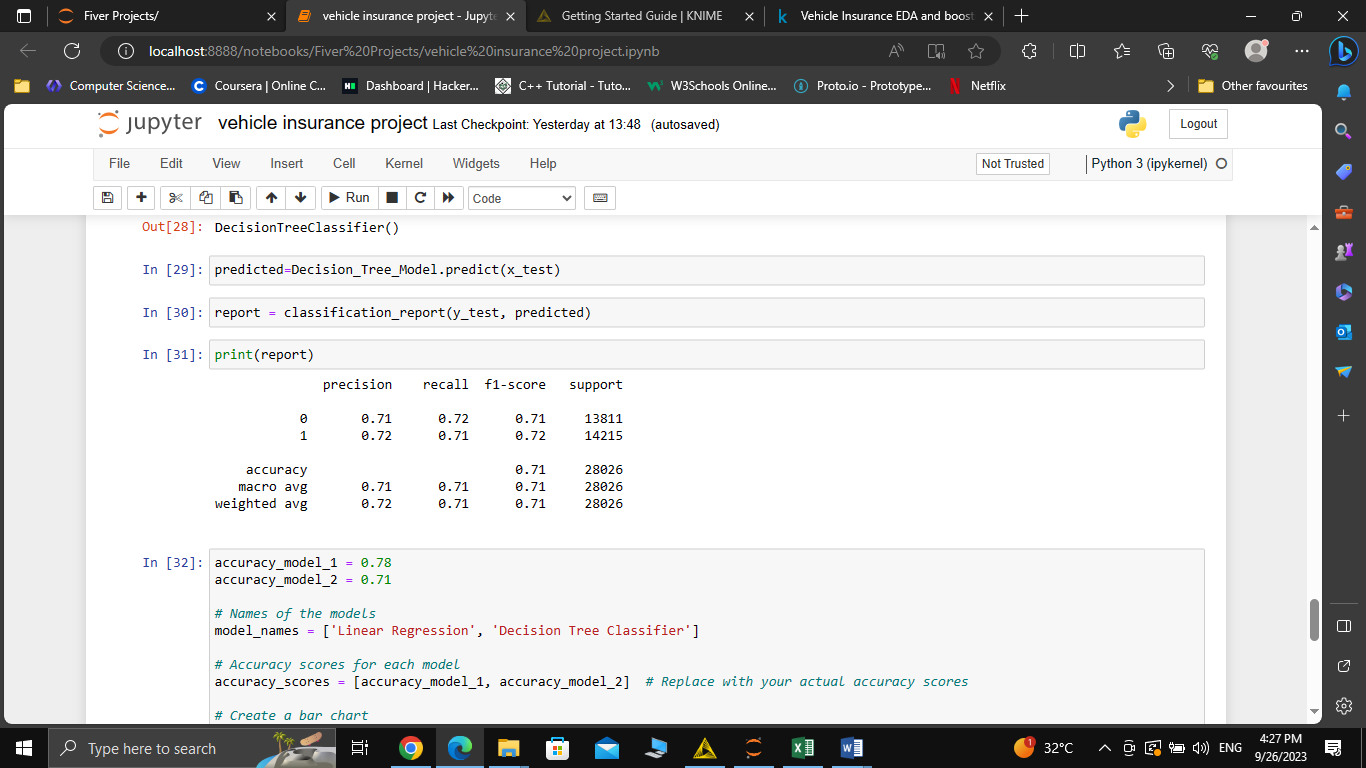
We employed two machine learning algorithms, logistic regression and decision tree classification, to construct predictive models. The dataset was partitioned into training and testing sets for model evaluation.

Model performance was assessed based on key metrics, including accuracy, precision, recall, and F1-score.

**Experimental Results**

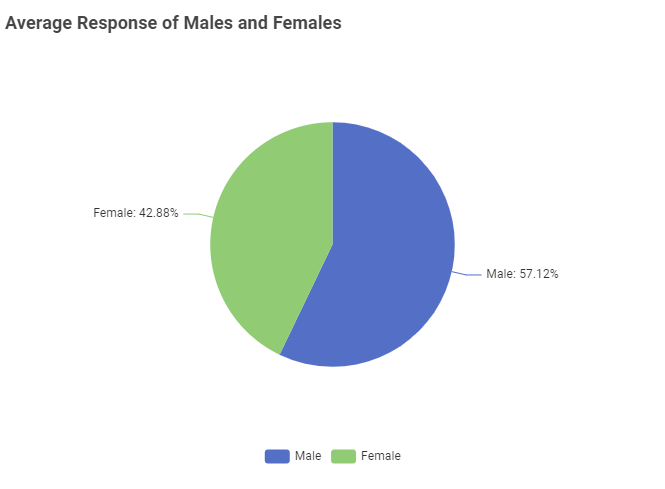
Linear regression Model Performance

Decision Tree Model Performance

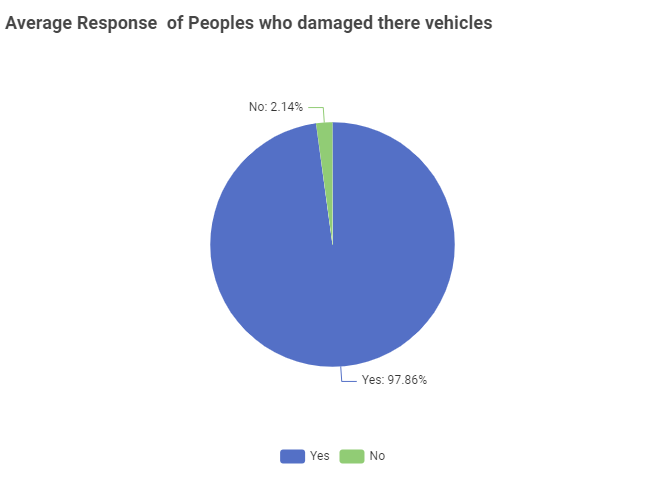
****

**Data Mining with KNIME:**

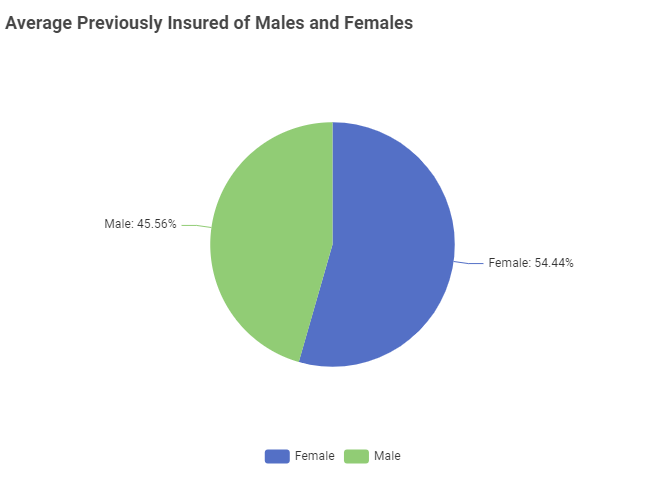
In addition to traditional modeling, we conducted data mining using KNIME, a data analytics and visualization tool. Visualizations such as pie charts and bar graphs were generated to gain insights into customer characteristics.



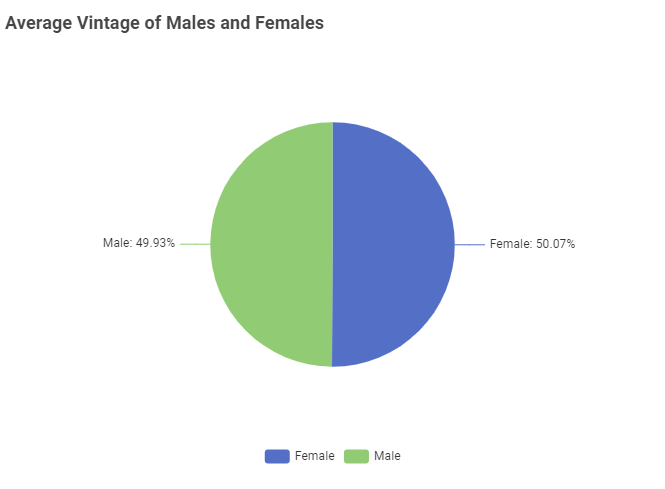
Here in the pie chart we can see that Males Response for showing interest in the insurance is greater than females



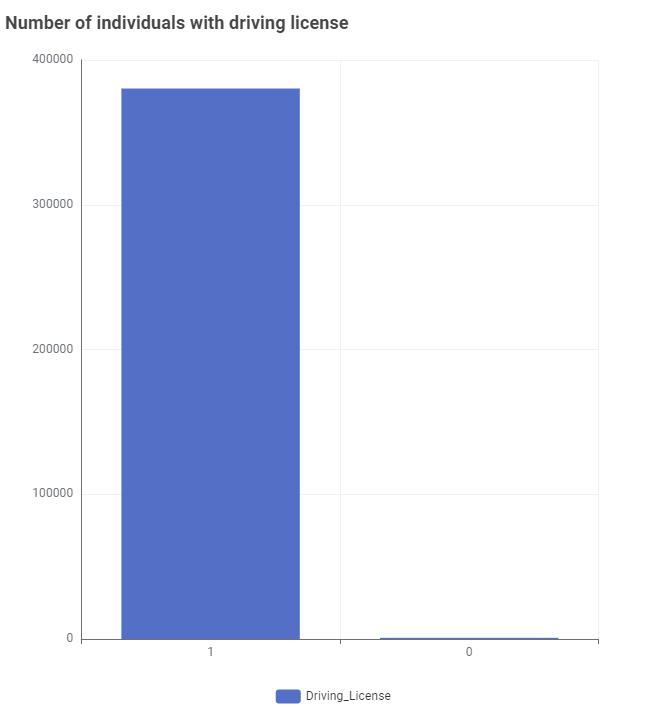
In this Pie chart Individuals who damaged their vehicles previously showing interest in Vehicle insurance more than who did not damaged their vehicles



Here in this pie chart show that females in this insurance company has more previously insurance than males



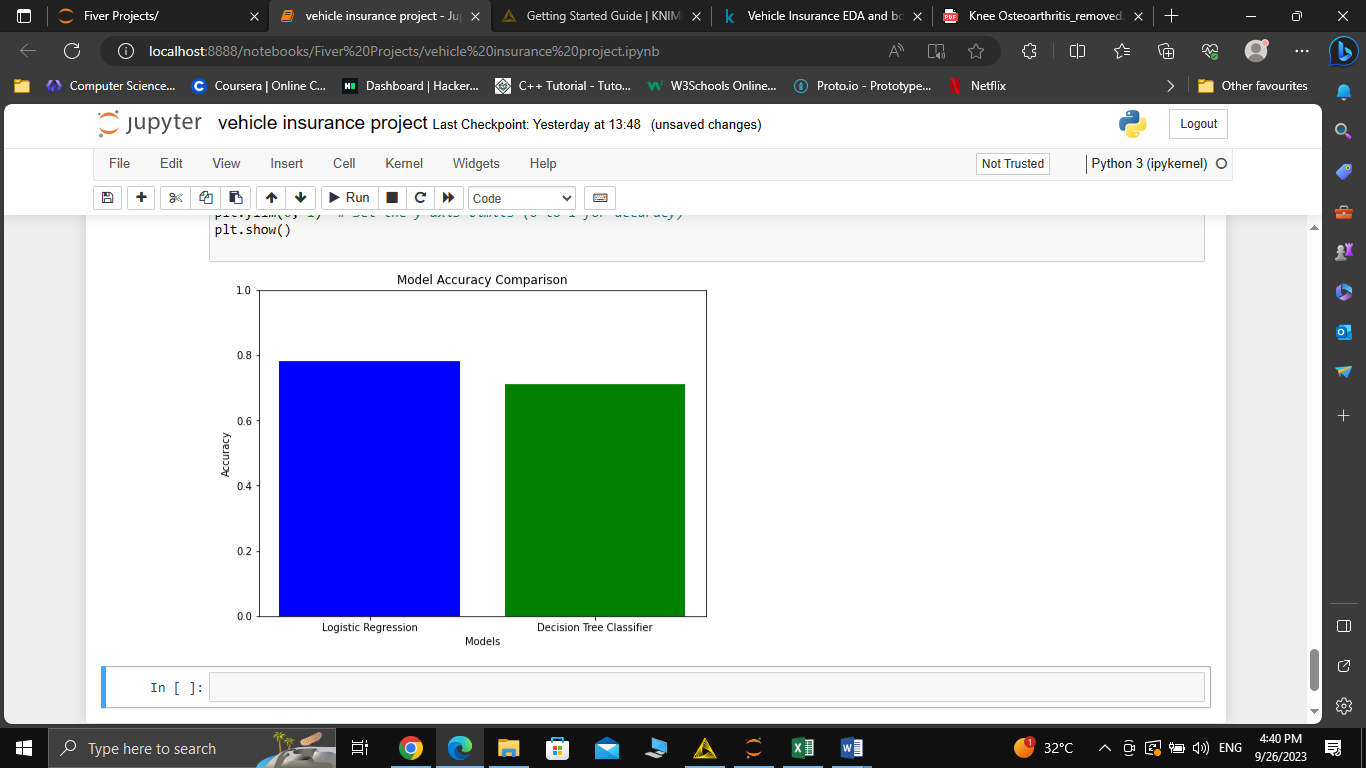
In this pie chart shows that Females has 50% Vintage Than Males

****

Here in the bar chart individuals with driving license is greater than who haven’t driving license

**Model Comparisons and Model Selection**

After training and evaluating both logistic regression and decision tree classification models, we compared their performance metrics. We can see in the figure Logistic Regression has better performance than Decision Tree Classifier. The model demonstrating the highest predictive accuracy was selected as the final model for identifying potential customers interested in vehicle insurance.



**Conclusions and Recommendations**

**Conclusions**

In conclusion, our analysis yielded the following insights:

* Successful Identification of Potential Customers: We successfully identified potential customers interested in vehicle insurance based on their health insurance policyholder status and other relevant attributes.
* Model Performance: The selected model demonstrated promising performance in predicting customer interest in vehicle insurance, as evidenced by its accuracy and other evaluation metrics.

**Recommendations**

Based on our findings, we propose the following recommendations:

* Implement the Predictive Model: Integrate the predictive model into the company's marketing and communication strategy to target potential customers effectively.
* Continuous Monitoring: Continuously monitor and evaluate the model's performance in real-world scenarios, and update it as necessary to maintain accuracy and relevance.
* Further Analyses: Conduct further analyses to refine the model and explore additional variables that may influence customer interest in vehicle insurance.

By implementing these recommendations, our client can capitalize on cross-selling opportunities, optimize revenue growth, and enhance the delivery of insurance services to their customers.

**Team Members:**

This project will be led by:

- [Your Name]

- [Team Member 2 Name (if applicable)]

This concludes the project report, outlining the comprehensive analysis and recommendations for our client's initiative to expand their insurance offerings.