**Project Report**

**Project Title**

**Cross-selling of Insurance Products using Predictive Analytics**

**Team Leader:**

N Sai Charan 21R11A0588

**Team Members:**

G . Sandhya Rani 21R11A0569

Dhathri 21R11A0545

Hasini 21R11A0518

Pooja 21R11A0501

PSBN Sriya 21R11A0595

MD.Akhbar Ali 21R11A0586

Shivanoor Vignesh 21R11A0597

**Exploratory Data Analysis**

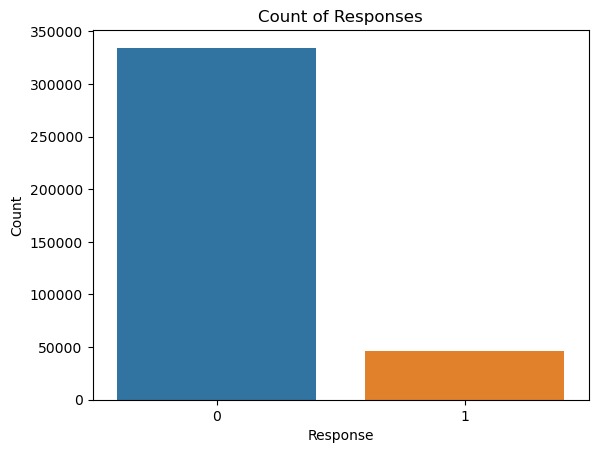
**Import data set**

For importing the dataset and to perform Exploratory Data Analysis we have to import some packages or library which are essential.

* import pandas
* import NumPy
* import seaborn

import matplotlib

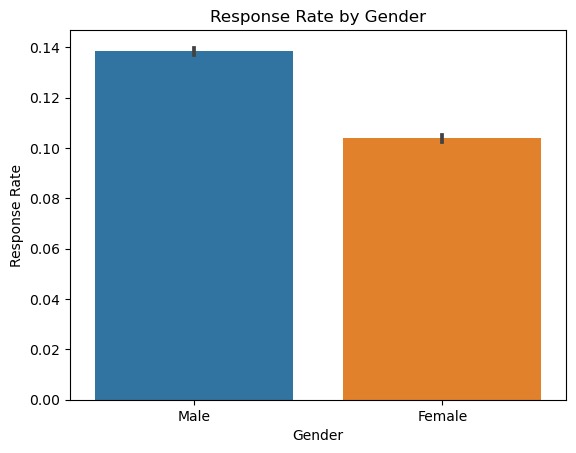
**1. Count of Responses**



**Observation:**

From the above bar plot, we can easily see that the majority of people are not interested in insurance, with very few indicating interest.

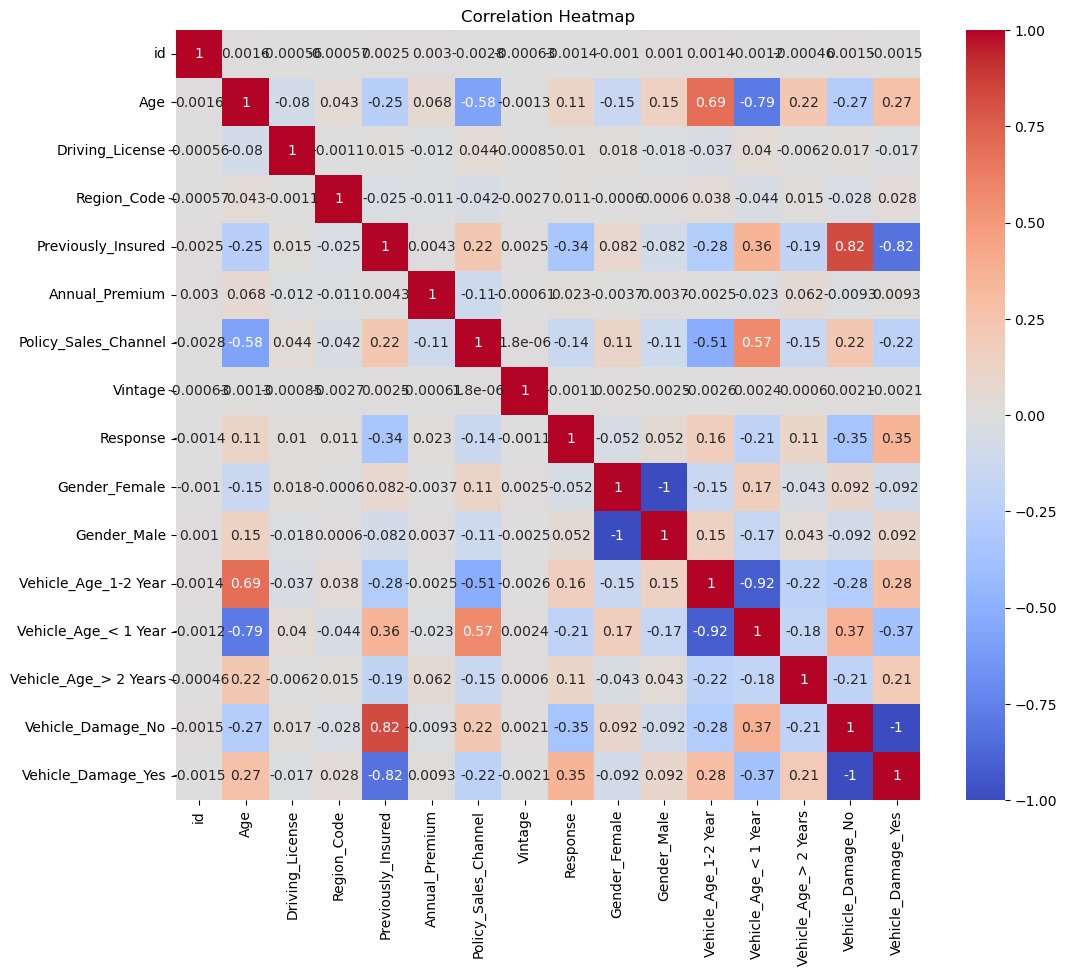
**2.Response Rate by Gender**



**Observation:**

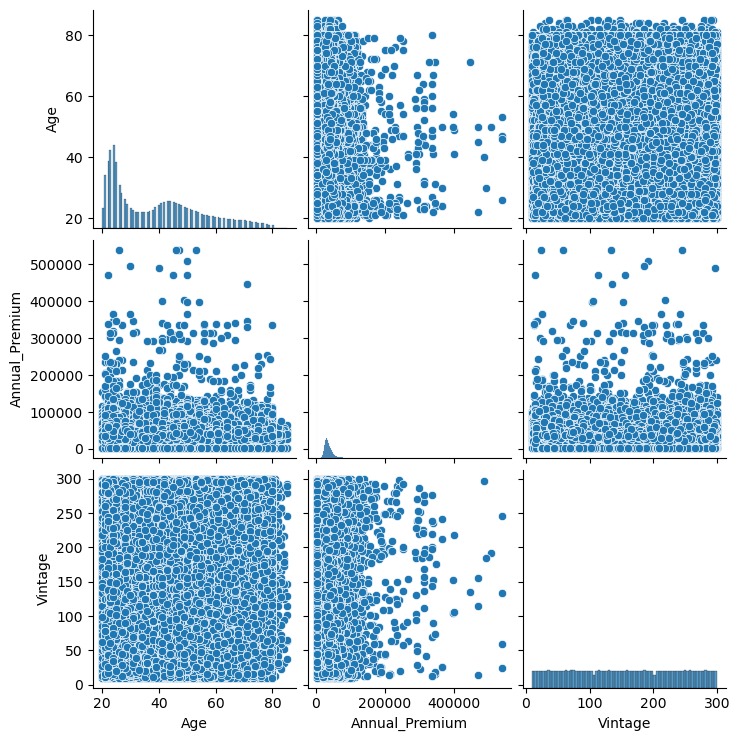
From the above graph, we can see that the response rates by gender differ slightly. The response rate for males is slightly higher than that for females.

**3.Heatmap**



**Observation**From the given heatmap, we can observe that age correlates positively with annual premium, and previously insured customers are less likely to respond positively. There are no strong negative correlations evident in the heatmap. The distribution of annual premium is heavily skewed towards lower values.

**4.Pairplot**



Observation

From the above pairplot , we can observe that the male response rate slightly exceeds the female response rate, with some variability indicated by error lines. The age distribution is skewed towards younger ages, suggesting a predominantly young population. In the plot of annual premium vs. age, there is a concentration of data points at younger ages and lower annual premiums. The vintage vs. age plot shows most data points clustered at lower ages, indicating newer policies among younger individuals. The distribution of annual premiums is heavily skewed towards lower values. In the plot of annual premium vs. vintage, there is no clear pattern, but many points are at lower annual premiums and vintages. The vintage distribution shows a fairly uniform distribution of vintages, without strong skew.

**Different Model prediction**

kNN Accuracy: 0.8565768413318989

Logistic Regression Accuracy: 0.8744850568077458

Decision Tree Accuracy: 0.8209572039568629

Random Forest Accuracy: 0.8647503345490803

The best model among the provided models is the decision tree, with an accuracy probability of 0.8209572039568629.

Based on the provided data and the output from the analysis of "Cross-selling of Insurance Products using Predictive Analytics,"

**Transaction 1:**

1:{ 14 items

"id": int 2

"Gender": string "Male" ✩

"Age" : int 76

"Driving License" : int 1

"Region\_code" : int 3

"Previously Insured" : int 0

"Vehicle\_Age" : string "1-2 Year"

"Vehicle\_Damage" : bool false

"Annual\_Premium" : int 33536

"Policy\_Sales\_Channel": int 26

"Vintage" int 183

"Response": int 0

"Scored Labels" : int 0

"Scored Probabilities" : float 0.029948773581753234

}

Observation:

* The individual’s ageof 76, lack of previous insurance, and low annual premium may contribute to the **low response probability** (scored probability of approximately 0.03).
* The **response received** for this transaction is **0**.
* This response value might indicate **denial** or a **non-affirmative response**.
* The above transaction vehicle age is 1-2years there is no previous insurance

**Transaction 2:**

2: {14 items

"id": int 3

"Gender": string "Male"

"Age" : int 47

"Driving License" : int 1

"Region\_Code" : int 28

"Previously Insured": int >

"Vehicle\_Age" : string "> 2 Years"

"Vehicle\_Damage" : bool true

"Annual\_Premium" : int 38294

"Policy\_Sales\_Channel" : int 26

"Vintage" int 27

"Response": int 1

"Scored Labels" : int 0

"Scored Probabilities" : float 0.3689280389686541

}

**Observation**

* The individual’s age, vehicle damage status, and higher annual premium may contribute to the **higher response probability** (scored probability of approximately 0.37).
* Being previously uninsured could also play a role in the positive response
* The **response received** for this transaction is **1**.
* This response value might indicate **approval** or a **positive response**

**Transaction 3**

3:{ 14 items

"id": int 4

"Gender": string "Male"

"Age" : int 21

"Driving License" : int 1

"Region\_code" : int 11

"Previously\_Insured" : int 1

"Vehicle\_Age" : string "< 1 Year"

"Vehicle\_Damage" : bool false

"Annual\_Premium" : int 28619

"Policy\_Sales\_Channel": int 152

"Vintage" : int 203

"Response" int 0

"Scored Labels" : int 0

"Scored Probabilities" float 0.00048811574341297133

}

**Observation**

* The individual’s age of 21, lack of previous insurance, and low annual premium may contribute to the **low response probability** (scored probability of approximately 0.0005).
* Being previously insured could impact the response outcome.
* The **response received** for this transaction is **0**.
* This response value might indicate **denial** or a **non-affirmative response**

**Transaction 4**

4:{14 items

"id": int 5

"Gender": string "Female"

"Age" : int 29

"Driving License": int 1

"Region\_Code": int 41

"Previously Insured" : int 1

"Vehicle\_Age" : string "< 1 Year"

"Vehicle\_Damage" bool false

"Annual\_Premium" : int 27496

"Policy\_Sales\_Channel" int 152

"Vintage" : int 39

"Response" int 0

"Scored Labels" : int 0

"Scored Probabilities" float 0.00035195474049872345

}

**Observation**

* The individual is **29 years old** and identified as **female**.
* They possess a **valid driving license**.
* The individual’s age, lack of previous insurance, and low annual premium may contribute to the **low response probability** (scored probability of approximately 0.0004).
* Being previously insured could impact the response outcome.
* The **response received** for this transaction is **0**.
* This response value might indicate **denial** or a **non-affirmative response**.

**Total Observations:**

* Transaction 2 has the highest predicted probability of 0.389, indicating a higher likelihood of interest in vehicle insurance. This transaction involves vehicle damage and a vehicle age of more than 2 years.
* Transactions 1, 3, and 4 have lower predicted probabilities because there is no vehicle damage in these transactions, and the vehicle age is almost below 2 years.