

# Travel Insurance Analysis

**Abstract—** Travel insurance is a fundamental part of the modern travel experience as it provides protection for travelers against various unforeseen events through their travels. This project delves into the looking and analyzing a dataset on travel insurance by utilizing data visualization and hypothesis testing techniques to uncover valuable insights that can enhance the understanding of travel insurance trends and potentially lead to better coverage options.

## I. INTRODUCTION

Travel insurance has become an integral aspect of modern-day travel, offering travelers a safety net against unexpected circumstances throughout their journeys. This project will employ data visualization and hypothesis testing techniques to find valuable insights that can enhance the understanding of travel insurance trends and potentially lead to more comprehensive coverage options. In a world marked by ever-evolving travel dynamics and unforeseeable challenges, delving into the world of travel insurance takes on a critical role in ensuring that individuals can traverse the globe with confidence and peace of mind.

## II. DATA PROCESSING AND VISUALIZATION

The dataset was found on Kaggle.com and is called 'Travel Insurance Prediction Data'. In the raw dataset there are ten features in the - Unnamed 0, Age, Employment Type, GraduateOrNot, AnnualIncome, FamilyMembers, ChronicDisease, FrequentFlyer, EverTravelledAbroad, and TravellInsurance.

### A. Data Processing

Several steps were taken to process and clean the raw dataset before visualizing the data. First, missing values were found, and their respective rows deleted. After the column named "Unnamed: 0" was dropped as it contains irrelevant information. Next, a function was created that transforms columns with 'Yes' and 'No' values into binary format, where 'Yes' is mapped to 1, and 'No' is mapped to 0. This function was applied to three specific columns: 'FrequentFlyer,' 'EverTravelledAbroad,' and 'GraduateOrNot,' effectively converting them into binary representations. Lastly, in the 'Employment Type' column for data consistency and simplification purposes the values 'Private Sector/Self Employed' were changed to 'Private Sector'.

### B. Exploratory Data Analysis and Visualization

Data visualization was employed to analyze the dataset, and from these visualizations, several trends and observations emerged, showcasing the characteristics and relationships within the data. In this dataset, the median age is 29, with the distribution exhibiting a positive skew, indicating that

there are more younger individuals than older individuals present. In addition, a correlation coefficient of -0.0201 was found - suggesting a very weak and almost negligible negative relationship between 'Age' and 'AnnualIncome.' This implies that changes in 'Age' are not strongly associated with changes in 'AnnualIncome' in this dataset and other factors likely play a more significant role in determining annual income.

Additionally, for annual income, those in the Private Sector generally have higher annual incomes compared to those in the Government Sector. The average income for employees in the Private Sector is approximately \$1,016,196, which is higher than the average income in the Government Sector which is approximately \$725,350. Furthermore, individuals in the Private Sector tend to purchase travel insurance at a higher rate than their counterparts in the Government Sector.

It was also found that the majority of people in this dataset have some form of chronic disease, are frequent flyers, have traveled abroad, and have travel insurance. In terms of age groups, more individuals in the '20-30' age group have purchased travel insurance compared to the '31-40' age group, however a higher number of those in the '20-30' age group do not have travel insurance. Additionally, more individuals in the 'Private Sector' purchased travel insurance compared to those in the 'Government Sector'. Graduates tend to have a higher annual income, and a higher proportion of graduates have purchased travel insurance compared to non-graduates. Moreover, individuals who purchased travel insurance tend to have a higher annual income. Interestingly, regardless of family size, most families choose not to purchase travel insurance.

## III. HYPOTHESIS TESTING

For the hypothesis tests, a significance level (alpha) of 0.05 was used. Other various metrics were calculated, but the primary metric used was the p-value for determining whether to reject or fail to reject the null hypothesis for each hypothesis.

### A. Hypothesis 1

The first hypothesis seeks to explore whether self-employed individuals tend to have higher incomes compared to those working in the private sector. For this test, the Null Hypothesis ( $H_0$ ) states that there is no significant difference in the average annual income between self-employed individuals and individuals working in the private sector. The Alternate Hypothesis ( $H_1$ ) states that self-employed individuals might indeed have a significantly higher average annual income when compared to their counterparts in the private sector. A p value of approximately  $1.31e-54$  was calculated, and it was found that private sector individuals have a significantly higher average annual income compared to those not in the government sector – meaning we rejected the null hypothesis. This visual can be seen in Figure 1.

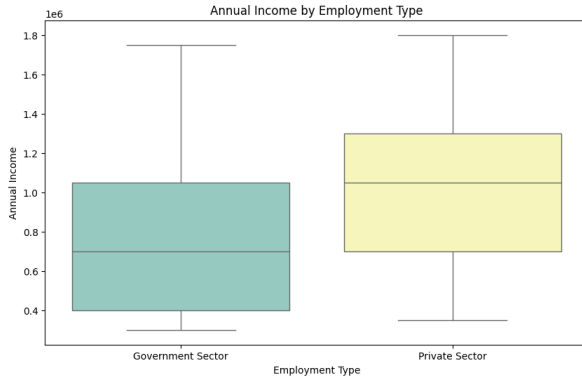


Fig. 1. Hypothesis 1

### B. Hypothesis 2

The second hypothesis looks at whether the number of family members influences the decision to purchase travel insurance. For this test, the Null Hypothesis (H0) is the average proportion of individuals who purchase travel insurance is the same across all different family sizes. The Alternate Hypothesis (H1) states that the average proportion of individuals who purchase travel insurance is not the same across all different family sizes. A p value of approximately 0.00048 was calculated, indicating that there is a relationship between the number of family members and travel insurance and that the null hypothesis is rejected.

### C. Hypothesis 3

The third hypothesis looks at whether individuals with chronic diseases are more inclined to buy travel insurance or not. The Null Hypothesis (H0) is the proportion of individuals with chronic diseases who purchase travel insurance is the same as the proportion of individuals without chronic diseases who purchase travel insurance. The Alternate Hypothesis (H1) is that proportion of individuals with chronic diseases who purchase travel insurance is different from the proportion of individuals without chronic diseases who purchase travel insurance. For this hypothesis, a p value 0.45 was calculated meaning there was a failure to reject the null hypothesis and that there is no association between chronic diseases and travel insurance.

### D. Hypothesis 4

The final hypothesis looks at if there is a significant difference in the proportion of people who have traveled abroad between those with and without travel insurance. The Null Hypothesis (H0) states the proportion of individuals who have traveled abroad is the same for those with travel insurance and those without travel insurance. The Alternate Hypothesis (H1) states that the proportion of individuals who have traveled abroad is different for those with travel insurance and those without travel insurance. After testing a p value of approximately  $1.41 \times 10^{-82}$  was found, indicating that the null hypothesis is rejected and that there is an association between traveling abroad and travel insurance. Looking at Figure 2 below, it shows that of those individuals who did travel abroad, the majority choose to purchase travel insurance.

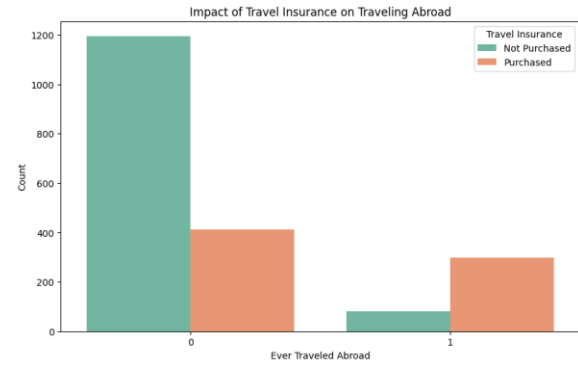


Fig. 2. Hypothesis 4

## IV. CONCLUSION

The results of our analysis are all shown Table 1 below.

TABLE I. HYPOTHESIS TESTING RESULTS

Hypothesis	Hypothesis Results
Employment vs Annual Income	Reject Null
Family Size vs Travel Insurance	Reject Null
Chronic Illness vs Travel Insurance	Do Not Reject Null
Travel Abroad vs Travel Insurance	Reject Null

Firstly, we found evidence indicating that non-government employee individuals have a significantly higher average annual income compared to those in the private sector. Secondly, our findings suggest that the number of family members does play a role in influencing the choice to acquire travel insurance. This highlights the impact of family size on insurance decisions. In addition, the analysis led to the conclusion that there is no substantial association between chronic diseases and travel insurance, as we failed to reject the null hypothesis in this case. Lastly, it was shown that there is a strong association between traveling abroad and the decision to purchase travel insurance. This highlights the significance of past travel experiences in shaping insurance choices. These results collectively provide valuable insights for both the insurance industry and potential travelers, enhancing our understanding of the factors influencing travel insurance decisions.