PORTO SEGURO SAFE DRIVER PREDICTION PROPOSAL

LEE YI

1.0 Introduction

Porto Seguro is Brazil's third biggest insurance firm, established in 1945. Porto Seguro focuses on the financial services business. This corporation offered a variety of financial services to customers, including auto insurance, health insurance, and life insurance. Furthermore, it is the market leader in Brazil's car and homeowner insurance divisions. To substantiate the statement above, it has around ten million customers across a variety of business lines, indicating that Porto Seguro is a reputable corporation. It demonstrates that Porto Seguro has a positive image with its consumers, since the majority of them purchase insurance from the company.

2.0 Business Goal

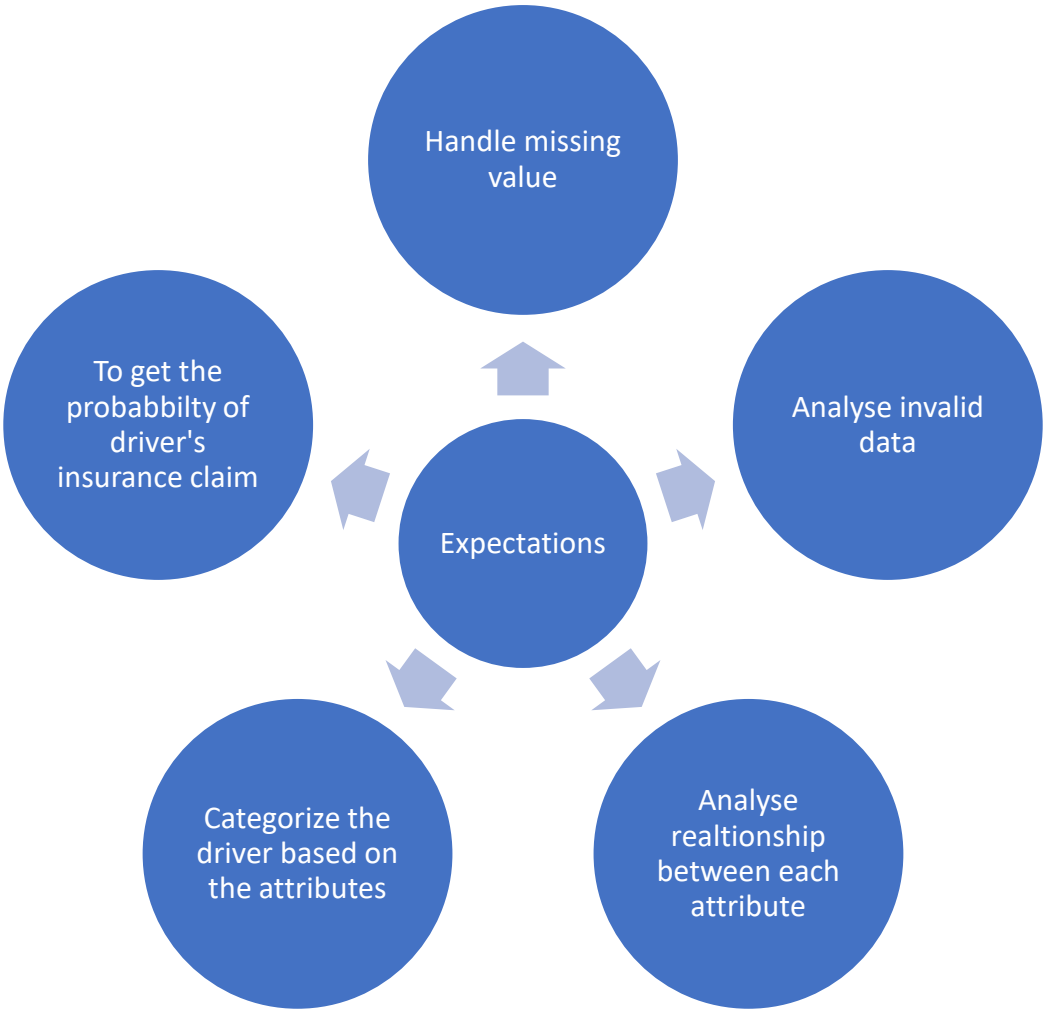
2.1 Company

For the last two decades, Porto Seguro has relied on machine learning. As a result of new government policies and economic growth, vehicle insurance has become a need in Brazil. Thus, this is an opportunity for Porto Seguro to differentiate itself and sell themselves to the public in order to get more individuals to purchase insurance from them. To find a better solution for pricing auto insurance, Porto Seguro launched a Kaggle competition to invite more professionals to develop a better machine learning model for estimating the chance of a motorist filing an insurance claim in the next year. Porto Seguro has a few objectives in mind with this assignment. To begin, the firm wants to investigate new and more sophisticated ways for making more accurate forecasts, which will enable it to further adjust its costs and make vehicle insurance coverage more affordable to a broader range of drivers from diverse backgrounds. Additionally, by having an error in the vehicle insurance company's claim forecasts, the cost of insurance for good drivers would climb, while the price for bad drivers will decrease. Therefore, it is critical to improve forecast accuracy in order for the corporation to set a more appropriate insurance premium. As a consequence, the business is able to retain existing clients while also attracting new ones. Additionally, if the firm offers reduced insurance rates for irresponsible drivers, the company's expenses would surge. Additionally, having a more accurate prediction model enables the business to expand its client base, particularly among low-income families.

2.2 Customer

Having a safe driver prediction algorithm helps not just the client, but also the present and prospective customer. To begin, clients may get more affordable insurance; the amount they must pay may be less than in the past. Thus, the money they saved may be used to assist alleviate the load or to obtain other forms of insurance for their family members.

3.0 Expectations



4.0 Outcome

Finally, but certainly not least, there are two possible solutions to this task. The objective of this study is to determine the best features that contribute to the basis for an insurance claim. As a result, the organization is able to distinguish between excellent and poor drivers and establish a suitable premium for insurance. Additionally, a machine learning model will be developed to provide a more accurate estimate of the driver's likelihood of filing an insurance claim.

根据下面要求，在右边写出对应代码	train	test
读取csv	`train = pd.read_csv('train.csv')`	`test = pd.read_csv('test.csv')`
行数	`train.shape[0]`	`test.shape[0]`
列数	`train.shape[1]`	`test.shape[1]`
共有多少元素	`len(set(train))`	`len(set(test))`
是否含有target	`True if 'target' in train.columns else False`	`True if 'target' in test.columns else False`
float64的变量有多少个?	`sum(train.dtypes == "float64")`	`sum(test.dtypes == "float64")`
int64的变量有多少个?	`sum(train.dtypes == "int64")`	`sum(test.dtypes == "int64")`
查看数据information	`train.info`	`test.info`
查看数据前3行	`traiin.head(3)`	`test.head(3)`
查看数据后8行	`train.tail(8)`	`test.tail(8)`
数据的列名	`train.columns`	`test.columns`
数据的维度	`train.ndim`	`test.ndim`
数据各列的数据类型	`train.dtypes`	`test.dtypes`
更改设置使数据最多只展示2行	`pd.set_option("display.max_rows", 2)`	`pd.set_option("display.max_rows", 2)`
更改设置使数据最多只展示5列	`pd.set_option("display.max_columns", 5)`	`pd.set_option("display.max_columns", 5)`