

Data Wrangling Process of Capstone Project 2

In this project, I work on the Auto Insurance claim records. Before starting to data wrangling, I explored my data. To be able to do that I applied shape, columns, dtypes, info attributes, and describe, isnull, duplicated methods of pandas to the data set.

After getting familiar with data, I find out three types of problems. These problems are:

- 1 Unnecessary information columns
- 2 Incompatible type of columns
- 3 White spaces in column names

There are outlier issues in this data set. I will handle them in data exploration.

Unnecessary information columns: At this stage, I filtered out unnecessary columns and dropped them. There were only 'country' column which has only 'usa' as a value. Therefore the information was not necessary..

```
df=df(columns=['column_name'])
```

Incompatible types of columns: Date column type was object. I changed it to datetime object.

```
df['column_name'] = pd.to_datetime(df['column_name'], errors='coerce')
```

White spaces in column names: There were some spaces between the word of column names. I removed these white spaces.

```
df.columns = df.columns.str.replace(' ', '')
```