Project: Car Insurance Claim

Overview:

- Dataset used in this project is of Car Insurance.
- The dimension of the dataset - 10302 customer's entries and 26 variables. Out of these entries we have 8737 distinct customers.
- For **Part 1 - Exploratory analysis and Visualization of Data** we have created 15 dashboards and for **Part 2 - Predictive Analytics and Visualisation of Insights** 4 dashboards.

Part 1 - Exploratory analysis and Visualization of Data:

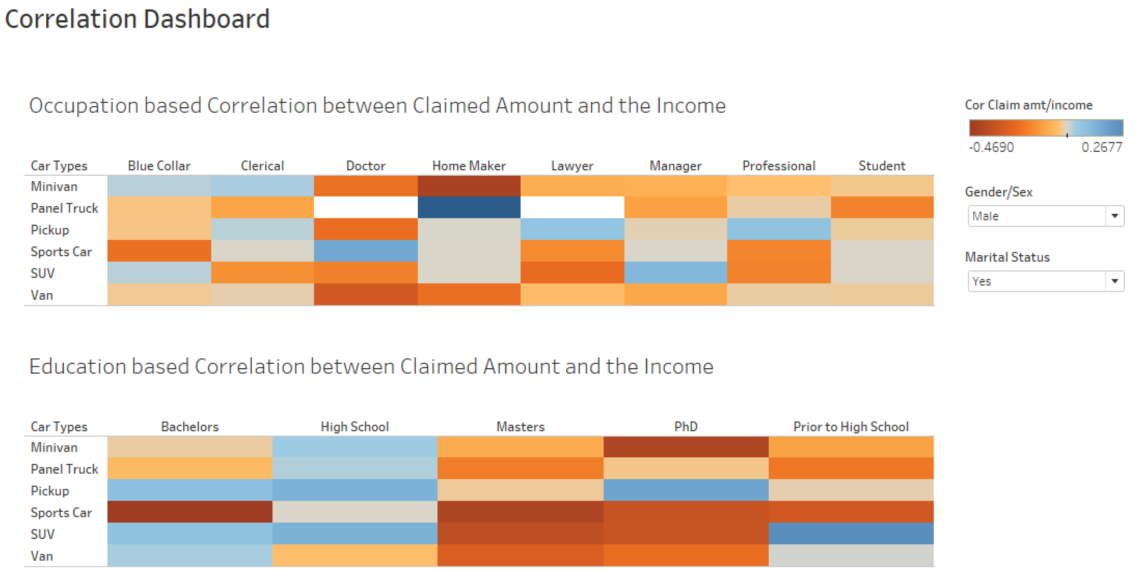
- **Record and Income Dashboard**: It's a multi-dimensional dashboard. We have plotted total number of customers and the average income of the customers falling into the multi-dimensional criteria. The dimensions in the multi-dimensions dropdown bar are – Education, Car Types, Car Use, Occupation, Gender/Sex and Marital Status. The selection from the dropdown changes the report accordingly.
  Insight – Customers with education PhD having the highest average income and Customers with occupation Blue Collar have the highest count.
- **Time Travel Dashboard:** Comparison of **a**verage travel time in Rural and Urban, travel time by private and commercial vehicle, Claim frequency and the travelled time.
  Insight – Time taken by private vehicles is more than the commercial ones. Also, average travel time in rural area is more in comparison with the urban one.
- **Claim Amount and Car Value:** Claim Amount and Car Value by gender and marital status.
  Insight – For Claim by gender and Marital status, average claim amount is claimed by the unmarried females. For Car value by gender and marital status – Minivan is having the highest car value among married females.
- **Car Type Usage in Rural and Urban:** Comparison of claim amount among car types and comparison of claims among rural/urban.
  Insight – Among Private car use, SUV holds the highest claim and in Commercial Pickup holds the highest claim. Among rural/ Highly-urban/Urban, Highly-urban/Urban is having the highest claims amount.
- **Bluebook Dashboard:** Comparison of Car value by level of Education and Occupation.
  Insights – For occupation, Value of car is highest among the Blue Collars and among education, Bachelors having the highest value.
- **Multi-dimensional Top N Claim and Frequency:** Multi-dimensional comparison of Claim amount and Claim frequency.
  Insight – Blue Collar holds the highest Claim amount and claim frequency, also PhD holder's claim amount and their frequency is minimum.
- **Claim by Education/Occupation with respect to Gender:**
  Insight – Among Blue collars highest claim frequency and the claim amount is highest among the males and female bachelors have the highest frequency.
- **Average Count based on Car type:** Comparison of customer count among different occupation and education.

Insight – Blue Collar (Occupation) and Bachelors (Education) have the highest customer count.

- **Distribution Dashboard:** Multi-dimension View of no. of records, income distribution and Claim amount distribution.
  Insight- Education: Prior to high school have the highest customer count, PhD having the highest income distribution, and the highest claim amount is claimed by High school customers.
  Occupation: Blue Collar have the highest customer count, Doctors having the highest income distribution, and the highest claim amount is claimed by Blue Collar customers.
- **Rank and Claim Dashboard:** Multi-dimensional Ranking based on claims. A chart of Claims based on the Age of Car (this is an interesting fact).
  Insight- Among Car types, Commercial, Pickup tops the ranking of claims and among Private it's SUV. Customer with 1 year of Car age have the highest claims.
- **Customer Info:** Comparison of no. of customers per Car Type, Customers Income per Car type vs No. of kids they are having, Average Income by education level and gender, car type and the Customer Age.
  Insight – SUV users are the highest and Panel truck users are minimum.
  Customers with Panel Truck and 0 kids have the maximum average income and Customers with Sports Car and 5 kids have the minimum average income. Average Income of PhD holder's among both gender is the highest and Prior to high school have the lowest average income. Customers with Minivan and year of Birth 1951 holds the maximum income.
- **Vehicle Info:** Blue book (It refers to the value of vehicle) comparison of car value per occupation, No. of customers using a particular Car type, Car value by Gender and marital status, Car value by rural/Urban.
  Insight – Blue Collar have the highest car value among occupation, Minivan holds the maximum value among private car types and Panel truck among commercial, and sports cars holds minimum car value among both private and commercial. Minivan among married females have the maximum car value and Sports car among married male is the minimum. Highly-urban have the highest car value.
- **Claim-Customers:** Frequency of claims as by gender, Claim amount by gender and marital status, Education vs Claim amount, Multi-Dimension frequency of claims.
  Insight – Claim frequency is greater for females. Unmarried females have the highest average claim amount. Bachelors have the maximum average claim amount. Blue collar have the highest claim frequency.
- **Claim by Car Info:** Claims by car type, Claim by occupation and Gender.
  Insight – Frequency of unique claims is highest for SUV. Blue collar male have the highest unique claim frequency and highest claim amount.
- **Claim-Aging History:** Comparison of Car Age and YOJ w.r.t claim amount.
  Insight- for YOJ- 12, we have the highest claim amount and for Car age 1 claim amount received is the highest.

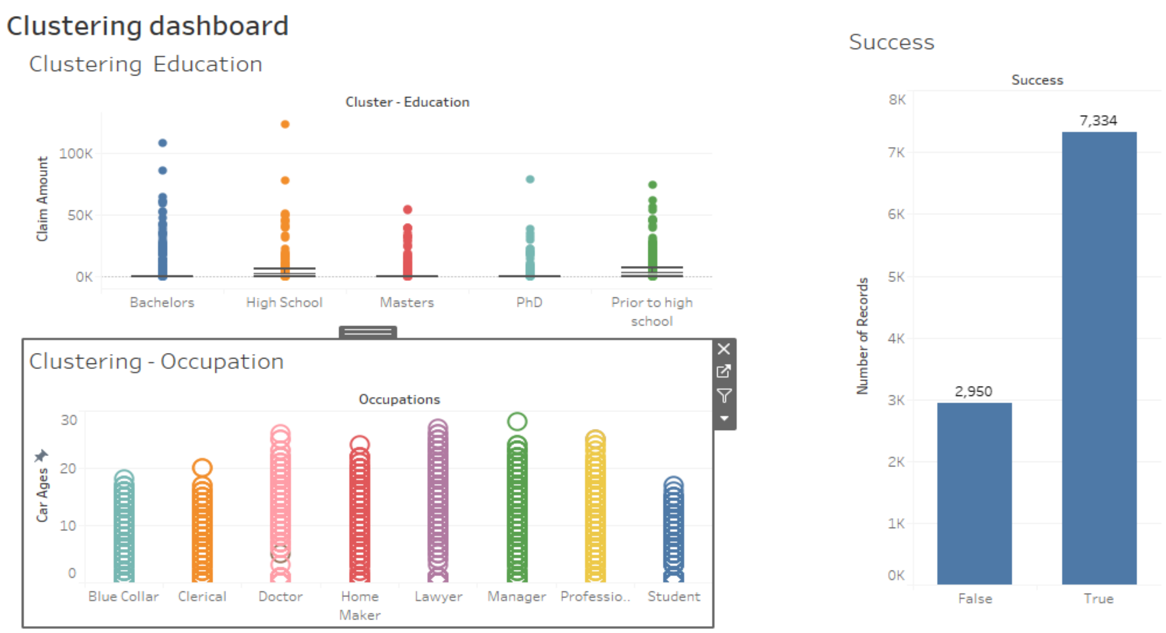
Part 2 - Predictive Analytics and Visualisation of Insights:

- **Correlation Dashboard:** Correlation between Claimed amount and the Income based on Occupation and the education. This correlation we are calculating for each Gender and their marital status.



Insight – A female manager who owns a Panel truck having the highest correlation (0.895) between the claim amount and the income. Similarly a married male Doctor who owns a sports car have the highest correlation between income and the claim amount.

- **Clustering dashboard:** We have created clusters based on education and occupation.



Insight –

Clustering Education - The average Claim amount is 1512.16, but as we can see there is an outlier in Education- High school with a claim amount of 123.25K. Success chart shows us that out of total of 10284, our cluster estimated 7334 correctly. ***Thus the accuracy of this cluster analysis is 71.31%.***

Summary

Count:	10284
SUM(Claim Amount)	
Sum:	15,551,013
Average:	1,512.16
Minimum:	0
Maximum:	123,247
Median:	0.00

Clustering Occupation –

Inputs for Clustering

Variables:	Occupations Avg. Income Avg. Claim Amount
Level of Detail:	Occupations
Scaling:	Normalised

Summary Diagnostics

Number of Clusters:	8
Number of Points:	9086
Between-group Sum of Squares:	8635.1
Within-group Sum of Squares:	80.499
Total Sum of Squares:	8715.6

Clusters	Number of Items	Centres		Most Common
		Avg. Income	Avg. Claim Amount	Occupations
Cluster 1	834	6267.9	1989.6	Student
Cluster 2	1501	33672.0	1585.8	Clerical
Cluster 3	785	12393.0	1379.9	Home Maker
Cluster 4	2165	58893.0	2089.5	Blue Collar
Cluster 5	1190	88161.0	698.69	Manager
Cluster 6	1323	75983.0	1535.1	Professional
Cluster 7	981	88381.0	1036.4	Lawyer
Cluster 8	307	1.2748e+05	585.21	Doctor
Not Clustered	533			

Analysis of Variance:

Variable	F-statistic	p-value	Model Sum of Squares	Model DF	Error Sum of Squares	Error DF
Avg. Income	632.1	0.0	64.23	7	131.8	9078
Avg. Claim Amount	14.38	0.0	0.1453	7	13.1	9078

Categorical variables are not included in the Analysis of Variance table.

- **Predicted Claim & Quadrant Analysis:** We have predicted the claim amount with respect to the car age.

  Predicted Claim*: The linear model script is:*

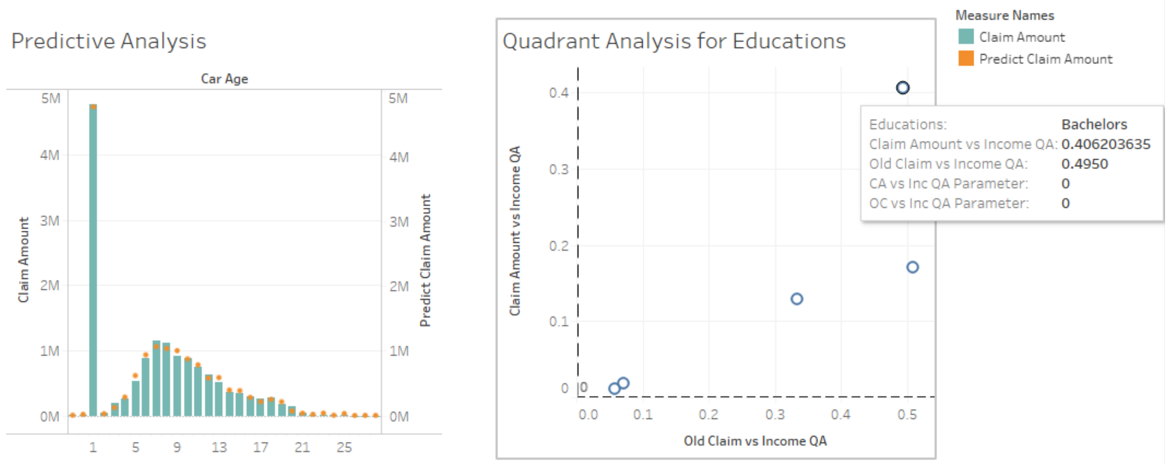
  Role: Continuous measure
  Type: Calculated Field
  Status: Valid

  SCRIPT_REAL("
  fit<- lm(.arg1 ~ .arg2 + .arg3 + .arg4 + .arg5)

  fit$fitted
  ",
  SUM([Claim Amount]),
  SUM([Car Ages]),
  SUM([Claim Frequency]),
  SUM([Kidsdriv]),
  SUM([Old Claim])
)

  Quadrant Analysis:

  Predicted Claim Amount w.r.t Car Age & Quadrant Analysis



  As we can see for bachelors, the correlation between the Claim amount vs Income is 0.40 and Old Claim vs Income is 0.49 and is the highest among all the education criteria.

- **Claim Frequency Time Series Forecasting: Forecasting using time-series modelling technique using exponential smoothing.**

Insights -

- ***With additive level, additive Trend and No Seasonality***

Options Used to Create Forecasts

Time series:	Car Ages
Measures:	Avg. Claim Amount
Forecast forward:	5 periods (28 – 32)
Forecast based on:	0 – 27
Ignore last:	1 period (28)
Seasonal pattern:	None (Searched for a seasonal pattern recurring every 1 Periods)

Avg. Claim Amount

Initial	Change From Initial	Seasonal Effect		Contribution		Quality
28	28 – 32	High	Low	Trend	Season	
432 ± 743	-221	None		100.0%	0.0%	Poor

All forecasts were computed using exponential smoothing.

Avg. Claim Amount

Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Additive	Additive	None	379	272	0.82	24.0%	343	0.000	0.047	0.000

Insight - With additive level, additive Trend and No Seasonality, the average estimated claim amount at Age 29 is 376. The MAPE (mean absolute percentage Error is 24% and AIC is 343.). Thus the accuracy is 76.0%.

- ***With additive level, additive Trend and additive Seasonality***

Options Used to Create Forecasts

Time series:	Car Ages
Measures:	Avg. Claim Amount
Forecast forward:	5 periods (28 – 32)
Forecast based on:	0 – 27
Ignore last:	1 period (28)
Seasonal pattern:	2 period cycle

Avg. Claim Amount

Initial	Change From Initial	Seasonal Effect		Contribution		Quality
28	28 – 32	High	Low	Trend	Season	
168 ± 875	-261	31 503	32 289	43.5%	56.5%	Poor

All forecasts were computed using exponential smoothing.

Avg. Claim Amount

Model			Quality Metrics					Smoothing Coefficients		
Level	Trend	Season	RMSE	MAE	MASE	MAPE	AIC	Alpha	Beta	Gamma
Additive	Additive	Additive	447	347	1.04	27.4%	356	0.000	0.338	0.255

Insight - With additive level, additive Trend and additive Seasonality, the average estimated claim amount at Age 29 is 317. The MAPE (mean absolute percentage Error is 27.4% and AIC is 356. Thus the accuracy is 72.6%.

Conclusion – We will consider the model with less MAPE and lower AIC value. Thus we will go for additive trend and no seasonality model.