# Car Insurance Claims

ERIC FOXCROFT

# Overview

- About the Stakeholder

- Business problem

- Context and Scope of Data

- Approach and Model compiled

- Visualizations

- Results / Recommendations / Conclusions

# About the Stakeholder …

- Car Insurance Company with various short term insurance products offered to a broad spectrum of Clients in the market …

- Engagement will focus on Representatives and Teams in the Company – focusing on product development, marketing and claims administration of their Car Insurance product …

# The Business problem …

- The Company needs to derive more value from available insurance & claims related data – to assist with:

    - New product development

    - Re-focus marketing initiatives

    - Optimize current Claims / Administration processes

" The goal is to analyse and predict **Client behavior** – in order to determine if a Client will claim or not claim against a Car Insurance Product ... "

# Context and Scope of Data …

Data set contains 10,000 records …

Details Client & Insurance related events with 19 features in the set …

Data granularity = each record represent a client & insurance product related activity …

# Data dictionary …

ID
AGE
GENDER
RACE
DRIVING EXPERIENCE
EDUCATION
INCOME
CREDIT SCORE
VEHICLE OWNERSHIP
VEHICLE YEAR
MARRIED
CHILDREN
POSTAL CODE
ANNUAL MILEAGE
VEHICLE TYPE
SPEEDING VIOLATIONS
DUIS
PAST ACCIDENTS
OUTCOME

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ID | AGE | GENDER | RACE | DRIVING_EXPERIENCE | EDUCATION | INCOME | CREDIT_SCORE | VEHICLE_OWNERSHIP | VEHICLE_YEAR | MARRIED | CHILDREN | POSTAL_CODE | ANNUAL_MILEAGE | VEHICLE_TYPE | SPEEDING_VIOLATIONS | DUIS | PAST_ACCIDENTS | OUTCOME |
| | 569520 | 65+ | female | majority | 0-9y | high school | upper class | 0.629027314 | 1 | after 2015 | 0 | 1 | 10238 | 12000 | sedan | 0 | 0 | 0 | 0 |
| | 750365 | 16-25 | male | majority | 0-9y | none | poverty | 0.357757117 | 0 | before 2015 | 0 | 0 | 10238 | 16000 | sedan | 0 | 0 | 0 | 1 |
| | 199901 | 16-25 | female | majority | 0-9y | high school | working class | 0.493145785 | 1 | before 2015 | 0 | 0 | 10238 | 11000 | sedan | 0 | 0 | 0 | 0 |
| | 478866 | 16-25 | male | majority | 0-9y | university | working class | 0.206012851 | 1 | before 2015 | 0 | 1 | 32765 | 11000 | sedan | 0 | 0 | 0 | 0 |
| | 731664 | 26-39 | male | majority | 10-19y | none | working class | 0.388365888 | 1 | before 2015 | 0 | 0 | 32765 | 12000 | sedan | 2 | 0 | 1 | 1 |
| | 877557 | 40-64 | female | majority | 20-29y | high school | upper class | 0.619127373 | 1 | after 2015 | 0 | 1 | 10238 | 13000 | sedan | 3 | 0 | 3 | 0 |
| | 930134 | 65+ | male | majority | 30y+ | high school | upper class | 0.49294355 | 0 | after 2015 | 1 | 1 | 10238 | 13000 | sedan | 7 | 0 | 3 | 0 |
| | 461006 | 26-39 | female | majority | 0-9y | university | working class | 0.468689297 | 0 | after 2015 | 0 | 1 | 10238 | 14000 | sedan | 0 | 0 | 0 | 1 |
| | 68366 | 40-64 | female | majority | 20-29y | university | working class | 0.521814936 | 0 | before 2015 | 1 | 0 | 10238 | 13000 | sedan | 0 | 0 | 0 | 0 |
| | 445911 | 40-64 | female | majority | 0-9y | high school | upper class | 0.561531032 | 1 | before 2015 | 0 | 1 | 32765 | 11000 | sedan | 0 | 0 | 0 | 1 |
| | 275820 | 65+ | male | majority | 30y+ | high school | upper class | 0.620361264 | 1 | after 2015 | 1 | 1 | 10238 | 10000 | sedan | 6 | 2 | 7 | 0 |
| | 521399 | 65+ | female | majority | 30y+ | high school | upper class | 0.729830882 | 1 | after 2015 | 1 | 0 | 32765 | 12000 | sedan | 4 | 0 | 0 | 0 |
| | 429728 | 40-64 | male | majority | 20-29y | high school | upper class | 0.637044744 | 1 | before 2015 | 1 | 1 | 10238 | 8000 | sedan | 4 | 1 | 2 | 0 |
| | 569640 | 16-25 | male | majority | 0-9y | university | upper class | 0.591259972 | 1 | before 2015 | 0 | 1 | 10238 | | sedan | 0 | 0 | 0 | 0 |
| | 980181 | 26-39 | male | majority | 10-19y | high school | middle class | 0.461567963 | 1 | before 2015 | 1 | 1 | 10238 | 12000 | sedan | 0 | 2 | 1 | 0 |
| | 906223 | 26-39 | female | majority | 0-9y | high school | upper class | 0.762797941 | 0 | after 2015 | 1 | 0 | 10238 | | sedan | 0 | 0 | 0 | 0 |
| | 517747 | 65+ | male | majority | 30y+ | university | upper class | 0.796174846 | 1 | before 2015 | 1 | 1 | 32765 | | sedan | 10 | 2 | 1 | 0 |
| | 24851 | 16-25 | male | majority | 0-9y | none | poverty | | 0 | before 2015 | 1 | 0 | 32765 | 12000 | sedan | 0 | 0 | 0 | 1 |
| | 104086 | 26-39 | female | majority | 0-9y | university | upper class | 0.68059428 | 1 | before 2015 | 0 | 1 | 32765 | | sedan | 0 | 0 | 0 | 1 |
| | 240658 | 16-25 | female | majority | 0-9y | high school | working class | 0.417713982 | 1 | before 2015 | 0 | 0 | 10238 | 18000 | sedan | 0 | 0 | 0 | 1 |
| | 484399 | 16-25 | male | majority | 0-9y | high school | working class | 0.409513531 | 0 | before 2015 | 0 | 0 | 10238 | 17000 | sedan | 0 | 0 | 0 | 1 |
| | 912828 | 16-25 | male | majority | 0-9y | high school | middle class | 0.553260053 | 1 | after 2015 | 1 | 1 | 32765 | 8000 | sedan | 0 | 0 | 0 | 1 |
| | 892754 | 40-64 | male | majority | 20-29y | none | poverty | 0.330950262 | 1 | before 2015 | 1 | 1 | 10238 | 13000 | sedan | 3 | 1 | 1 | 0 |

| | ID | CREDIT_SCORE | VEHICLE_OWNERSHIP | MARRIED | CHILDREN | POSTAL_CODE | ANNUAL_MILEAGE | SPEEDING_VIOLATIONS | DUIS | PAST_ACCIDENTS | OUTCOME |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 10000.000000 | 9018.000000 | 10000.000000 | 10000.000000 | 10000.000000 | 10000.000000 | 9043.000000 | 10000.000000 | 10000.00000 | 10000.000000 | 10000.000000 |
| mean | 500521.906800 | 0.515813 | 0.697000 | 0.498200 | 0.688800 | 19864.548400 | 11697.003207 | 1.482900 | 0.23920 | 1.056300 | 0.313300 |
| std | 290030.768758 | 0.137688 | 0.459578 | 0.500022 | 0.463008 | 18915.613855 | 2818.434528 | 2.241966 | 0.55499 | 1.652454 | 0.463858 |
| min | 101.000000 | 0.053358 | 0.000000 | 0.000000 | 0.000000 | 10238.000000 | 2000.000000 | 0.000000 | 0.00000 | 0.000000 | 0.000000 |
| 25% | 249638.500000 | 0.417191 | 0.000000 | 0.000000 | 0.000000 | 10238.000000 | 10000.000000 | 0.000000 | 0.00000 | 0.000000 | 0.000000 |
| 50% | 501777.000000 | 0.525033 | 1.000000 | 0.000000 | 1.000000 | 10238.000000 | 12000.000000 | 0.000000 | 0.00000 | 0.000000 | 0.000000 |
| 75% | 753974.500000 | 0.618312 | 1.000000 | 1.000000 | 1.000000 | 32765.000000 | 14000.000000 | 2.000000 | 0.00000 | 2.000000 | 1.000000 |
| max | 999976.000000 | 0.960819 | 1.000000 | 1.000000 | 1.000000 | 92101.000000 | 22000.000000 | 22.000000 | 6.00000 | 15.000000 | 1.000000 |

# Approach and Model compiled …

## Approach

- This output can be obtained via the use of a <u>Machine Learning (ML) model</u> - that will analyze and <u>predict</u> the potential "<u>Claim Status</u>" (target value) for each Client.
- Key steps:
  - Define the problem
  - Collect the data
  - Clean the data
  - Analyse the data
  - Choose and compile the appropriate ML model
- From a ML modelling perspective - this is a typical <u>Binary Classification</u> requirement to solve.

## Insurance Claims
## (Machine Learning Model)

- Model outputs a predicted "Target" value for Claim "Outcome" on each Client.
- This value represents the possibility that a Client will claim against their Insurance Policy.
- This will be based on analysis & learnings from the Customer's behaviour and data available to the Model.
- Outcome:
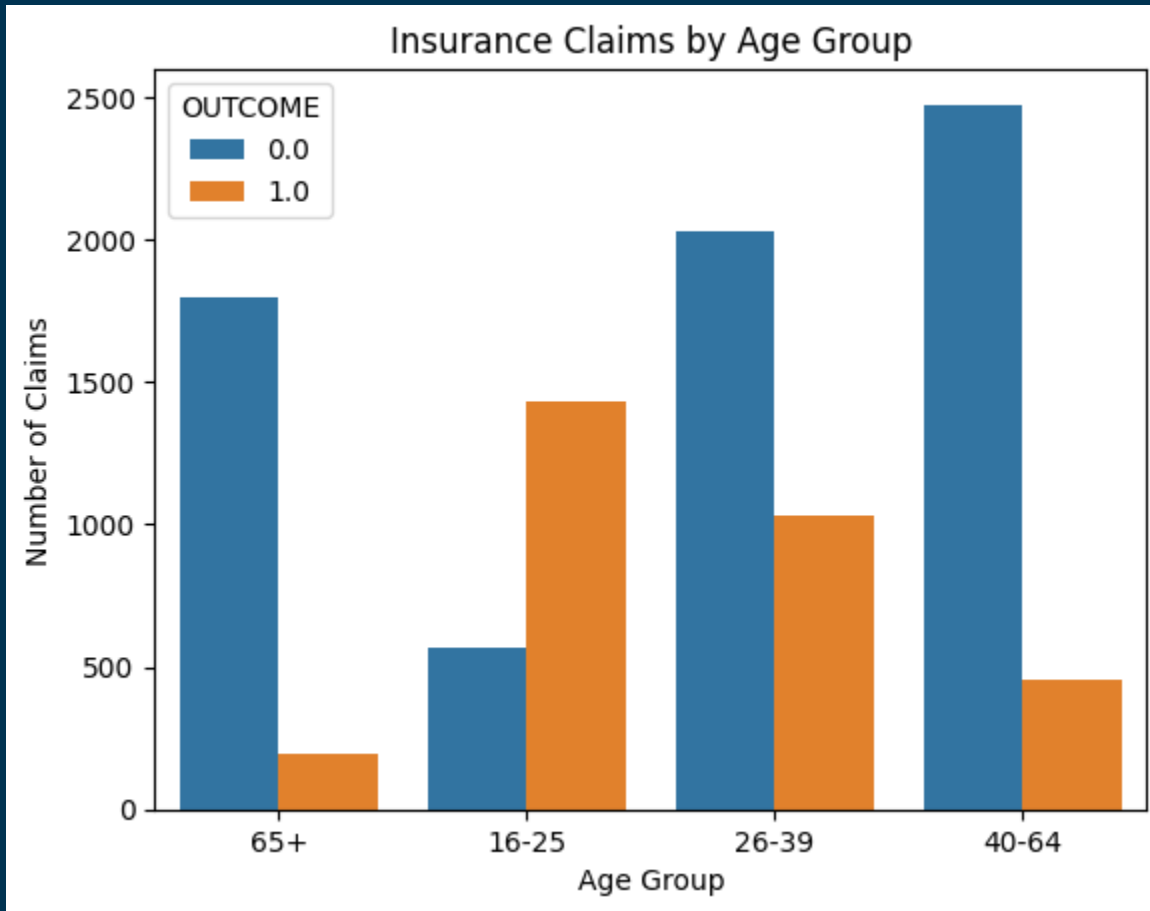  - 1 = Claimed
  - 0 = Not Claimed

# Visualizations …



Comparison: Driving Experience vs. Claims
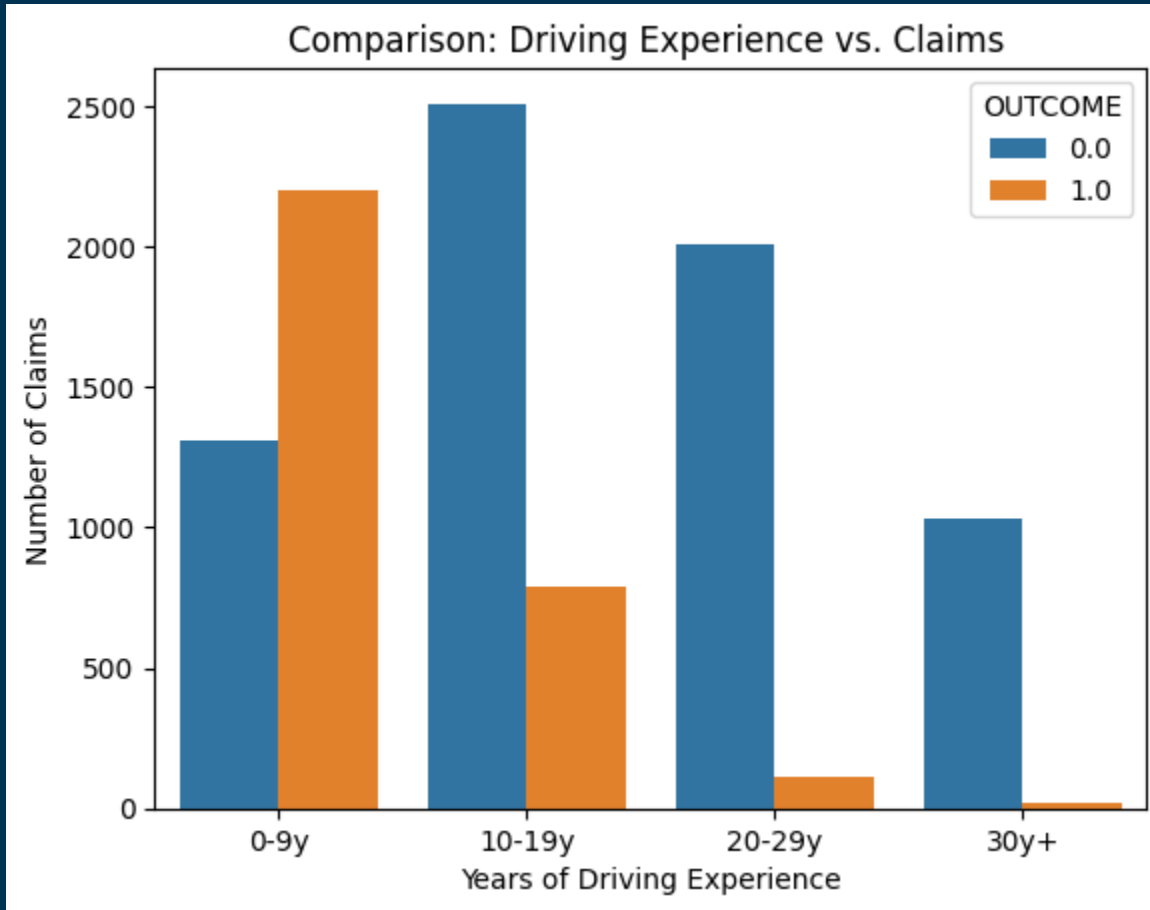
"Claims are <u>reduced</u> by Clients with more `Years of Driving Experience` …
e.g. 20-29 year & 30+ year groups

Claims are <u>more likely</u> to occur in the group where Clients have `0-9 Years of Driving Experience` … "

# Visualizations ...



"Insurance Claims are the <u>highest</u> in the `Age Group` 16-25 year and much lower in the 65+ year group ..."

# Results …

- 3 x ML Models Compiled
    - K-Nearest Neighbors (KNN)
    - Logistic Regression Model
    - Logistic Regression - using PCA

- Preferred model result …
    - **Logistic Regression Model - using Principal Component Analysis (PCA)***
    - Preference = based on Model accuracy rate
    - Model achieved an **accuracy rating of 86%**

*\* Model compiled by using Principal Component Analysis (PCA) - to reduce the dimensionality (number of features) within the dataset*

- Model outcomes – limited by the current available features and data elements from the data set.

- Possible to further increase the model accuracy by including additional elements if / when available.

# Recommendations / Conclusions ...

- Management of **claims & costing** of insurance products (premiums) can be optimized by taking note of the impact of some of the referred data elements in the Model outcomes and visualizations.

- Examples of some of these events that increase claim incidents are:

  - Clients with Speeding Drive Styles
  - Clients with limited years of Driving Experience
  - Clients that fall in younger age groups
  - Client with lower income levels, etc.

- The Insurance Company can also potentially **increase profits** (reduce claims) by expanding the below segments of their current Client base and **market share**:

  - Clients with 20+ years driving experience
  - Clients of 40+ years of age, etc.