TRAVEL INSURANCE CLAIM RATE

CURATED BY:ARIA DIPA ITNA ERLANGGA

OUTLINE

- Business Understanding
  - Problem Statements
  - Goals
  - Analytics Approach
  - Metric Evaluation
  - Data Information
- Exploratory Data Analys
  - Claim rate agency type
  - Net sales before and after data cleaning
  - Distribution of Claim and Commission as per Product Name
  - Distribution of Claim and Commission as per Agency
  - Relation between registered Claims and Product Name
  - Product vs Net Sales
  - Agencies with max number of Net Sales
- Model and Approaches
- Score on Vanilla Model
- Confusion Matrix
- Kesimpulan
- Rekomendasi

BUSINESS UNDERSTANDING

- Sebuah perusahaan yang bergerak di bidang Asuransi Perjalanan (Erlangga Travel Co., Ltd.) ingin mengetahui pemegang polis yang akan mengajukan klaim asuransi atas pertanggungan. Lebih dari itu perusahaan asuransi juga bertugas untuk menanggung resiko customer, sehingga risk management merupakan aspek penting dalam industri asuransi.
- Data pemegang polis di perusahaan asuransi merupakan data historis, terdiri dari destinasi, produk asuransi, dan sebagainya.

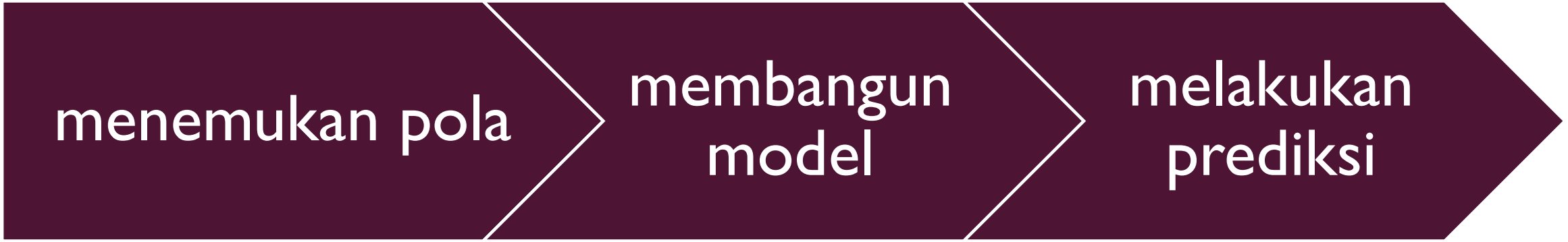
PROBLEM STATEMENT

- Apakah nasabah/customer akan melakukan claim atau tidak?

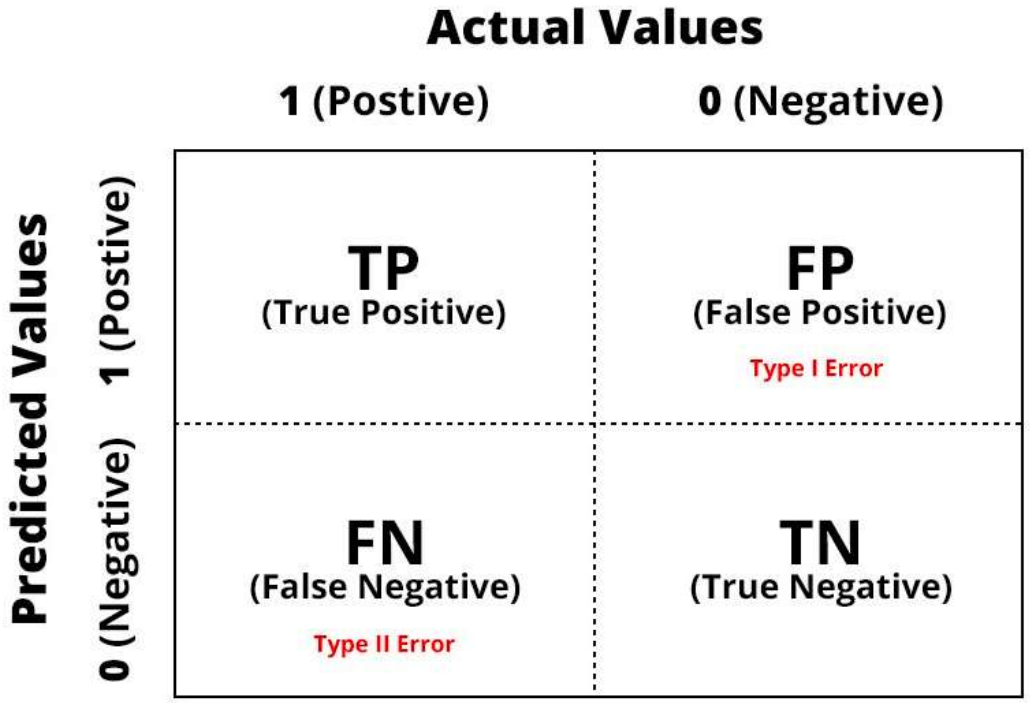
GOALS

- Memaksimalkan penjualan polis untuk product yang jarang di claim
- Melakukan prediksi terhadap customer akan claim atau tidak
- Meminimalisir tingkat False Negative (diprediksi tidak claim, ternyata claim)

ANALYTIC APPROACH



METRIC EVALUATION



Source: https://scikit-learn.org/stable/modules/generated/sklearn.metrics.recall_score.html?highlight=sklearn+metrics+recall_score

DATA INFORMATION

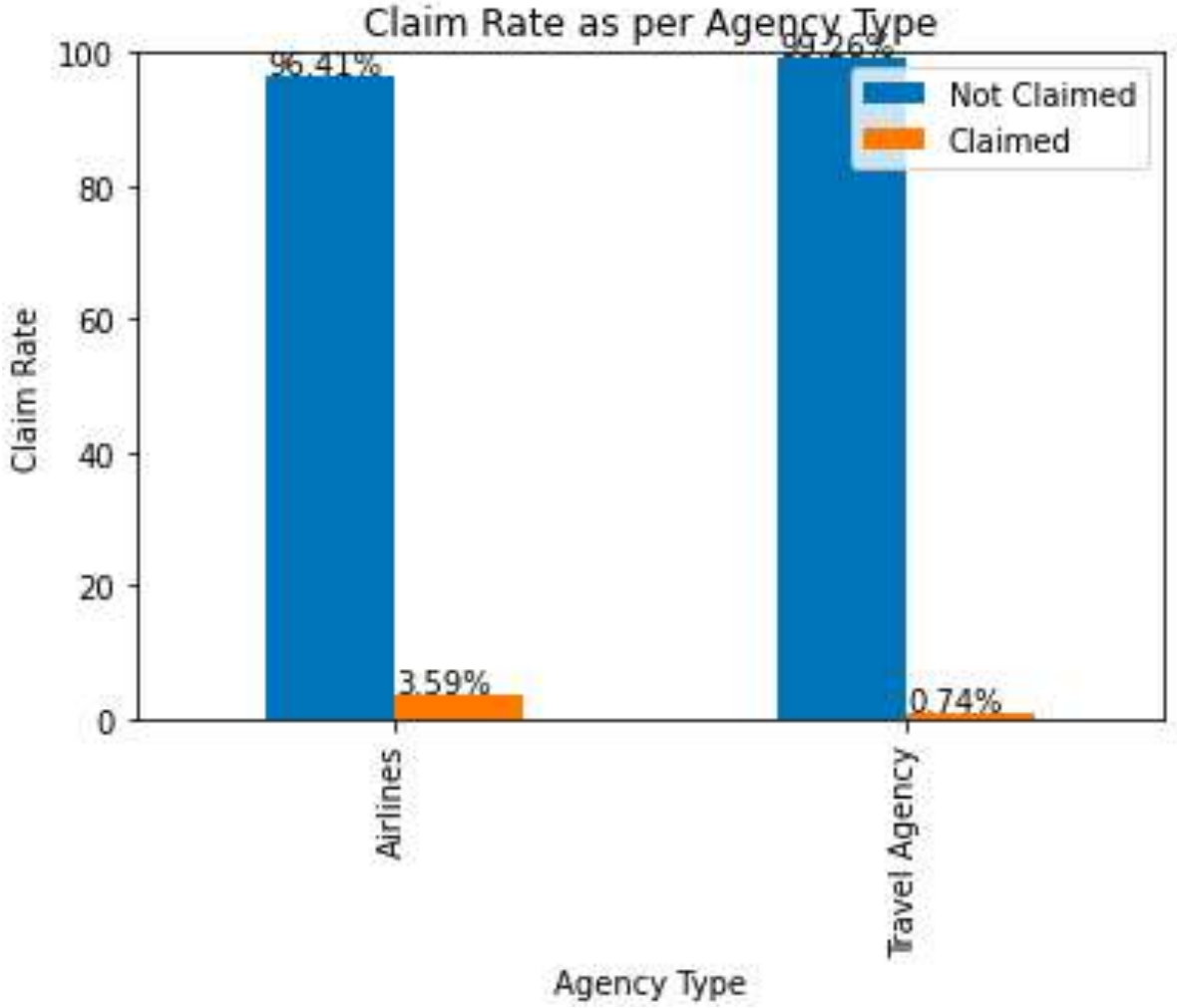
Feature	Datatype	Description
Agency	object	Name of agency
Agency Type	object	Type of travel insurance agencies
Distribution Channel	object	Distribution channel of travel insurance agencies
Product Name	object	Name of the travel insurance products
Duration	int64	Duration of travel
Destination	object	Destination of travel
Netsales	float64	Amount of sales of travel insurance policies
Commision (in value)	float64	The commission received for travel insurance agency
Age	int64	Age of insured (Age)

Target	Datatype	Description
Claim	int64	Claim status 1(yes) / 0(no)

EXPLORATORY DATA ANALYSIS

- Claim rate agency type
- Net sales before and after data cleaning
- Distribution of Claim and Commission as per Product Name
- Distribution of Claim and Commission as per Agency
- Relation between registered Claims and Product Name
- Product vs Net Sales
- Agencies with max number of Net Sales

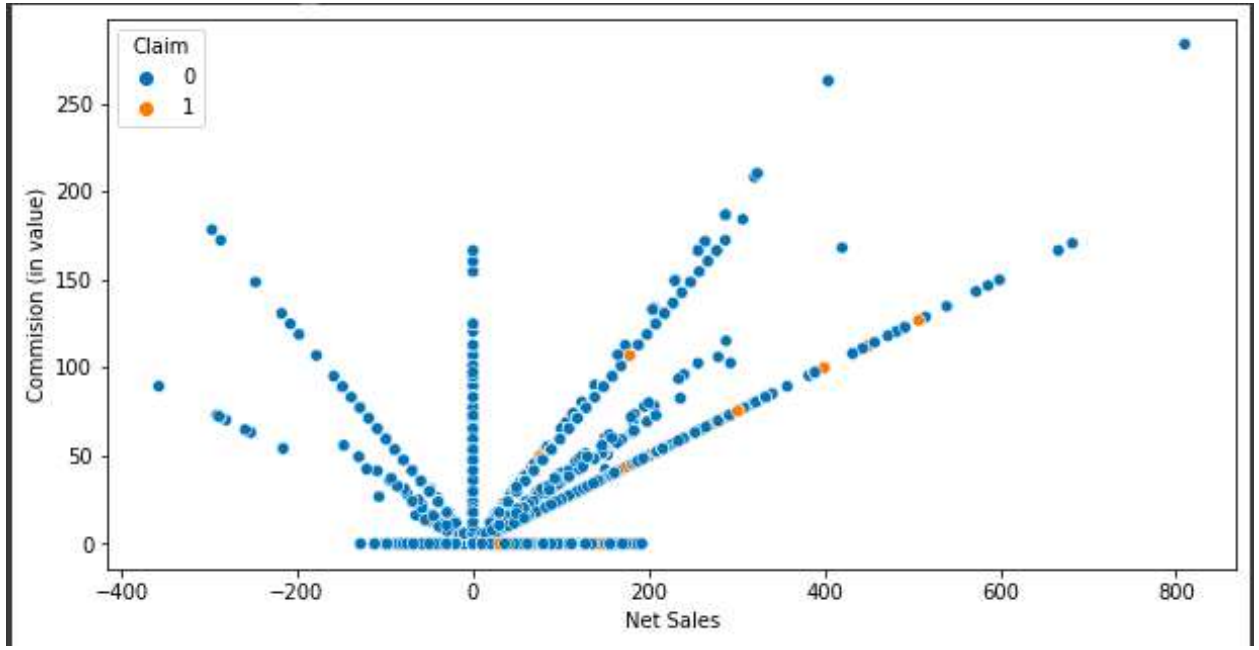
EDA - CLAIM RATE AGENCY TYPE



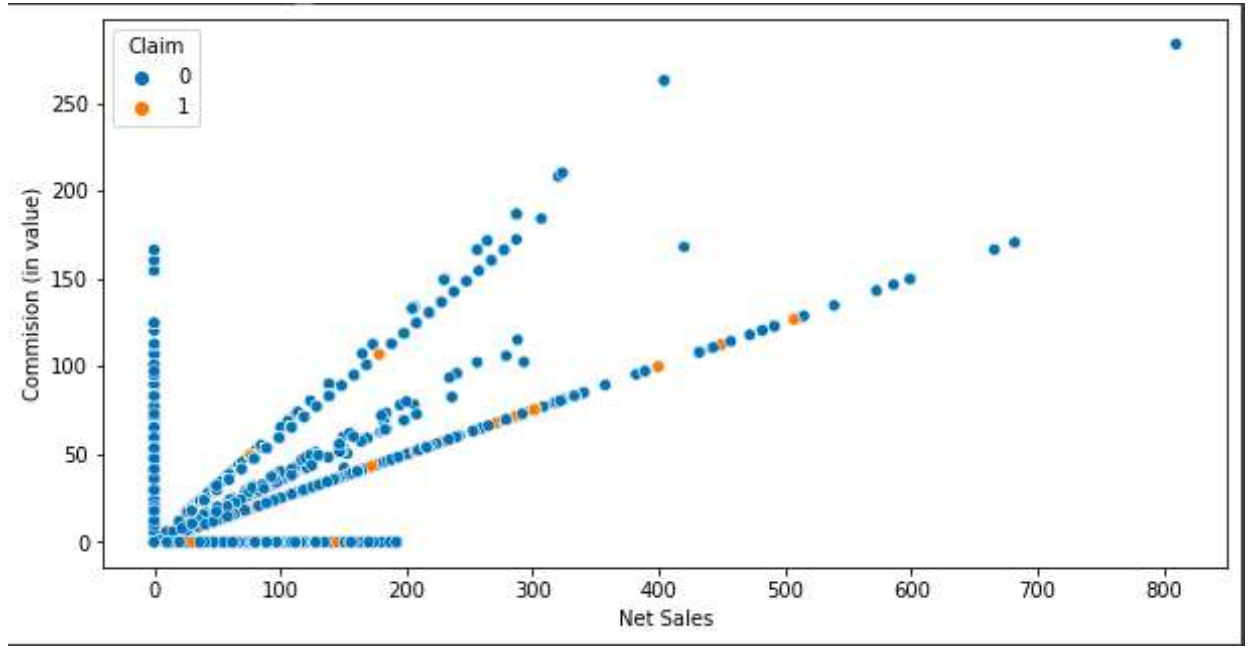
Class Imbalance: distribution of target variable (Claimed and Not Claimed) shows clear imbalance in to classes.

✓ SMOTE, ROS, RUS, Class Weight is used to balance distribution of target variable

EDA - NET SALES BEFORE AND AFTER DATA CLEANING

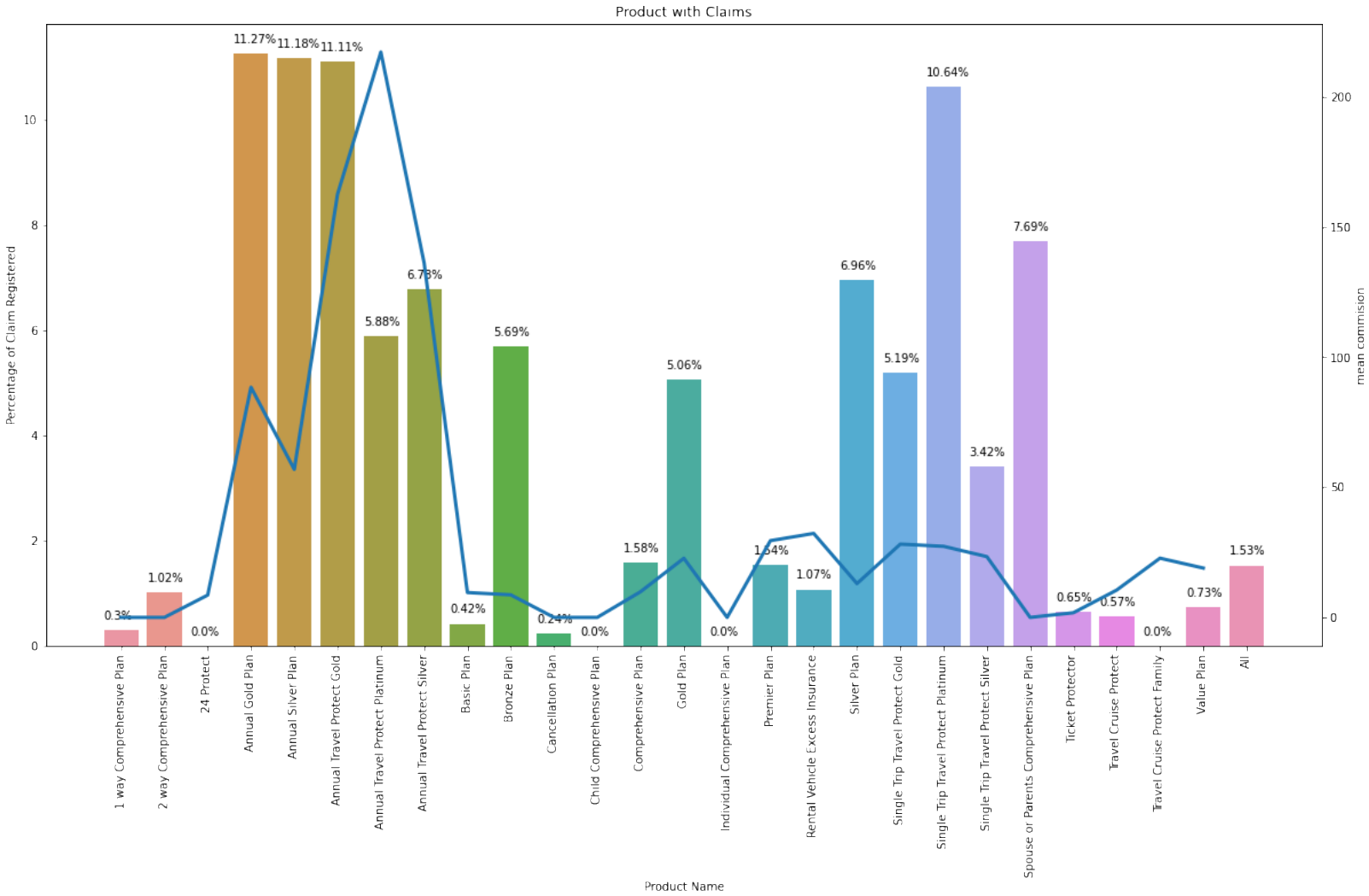


before



after

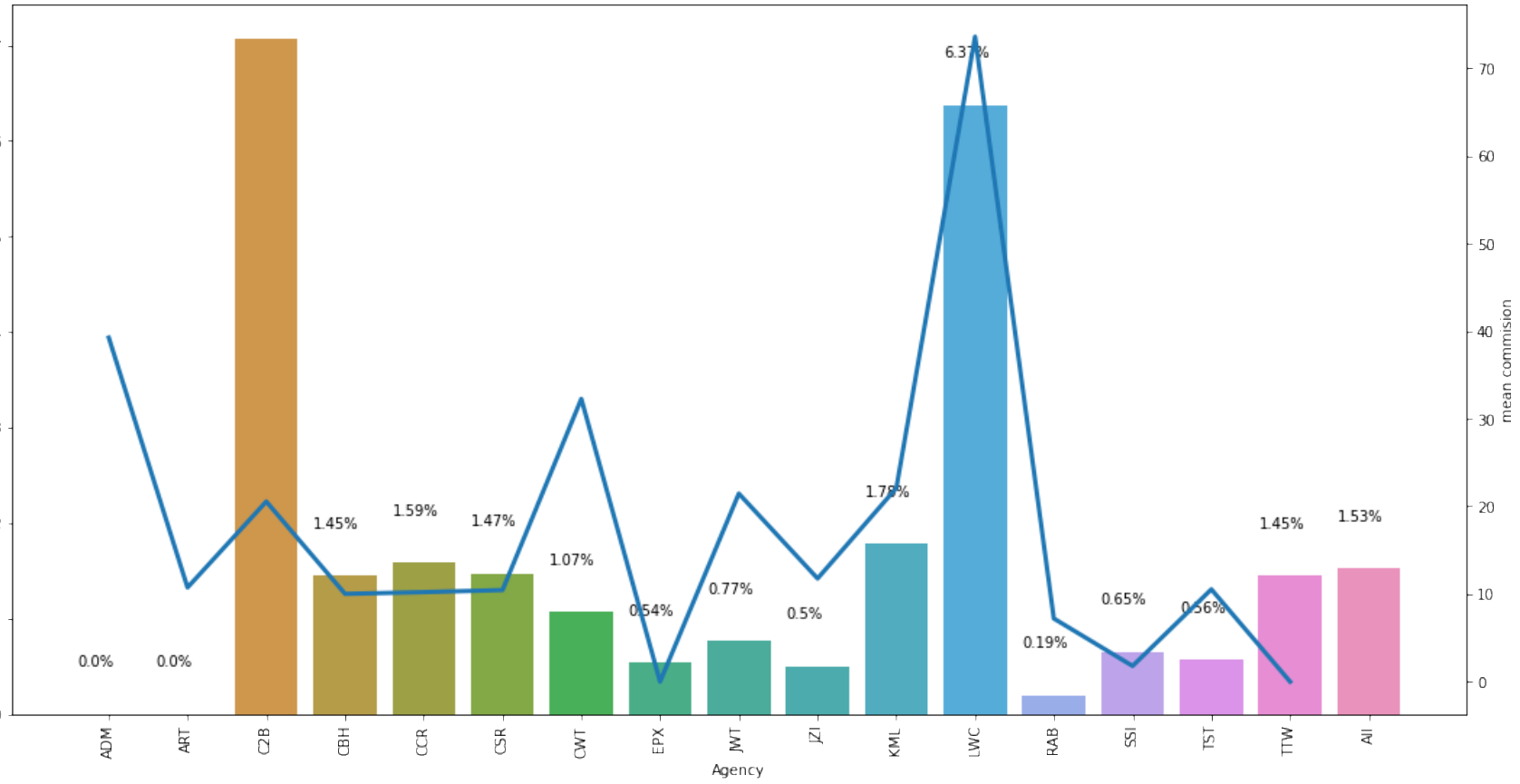
EDA - DISTRIBUTION OF CLAIM AND COMMISSION AS PER PRODUCT NAME



- Annual Gold & Annual Silver Plan have higher number of Claims registered whereas commission drawn is very low

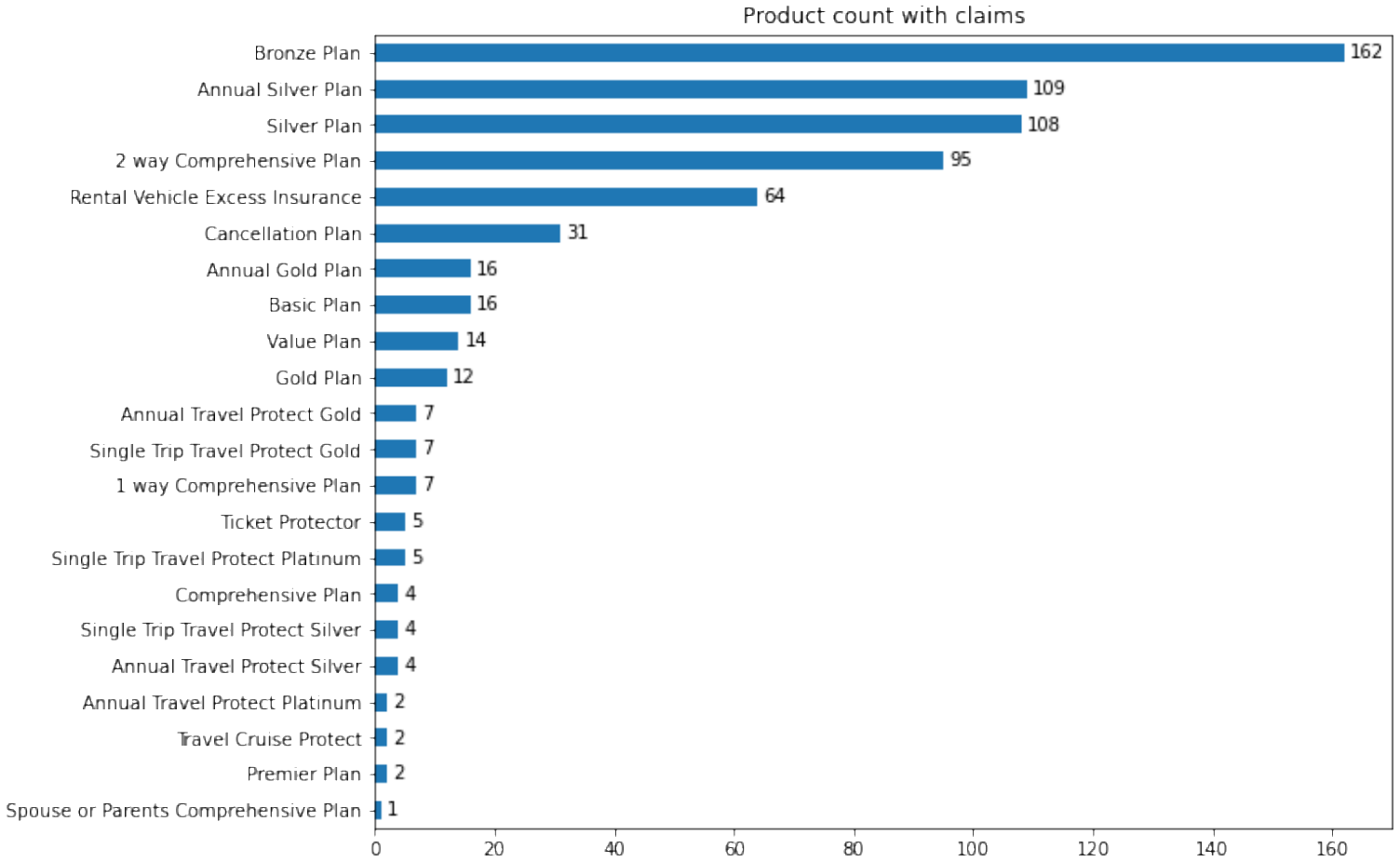
- Annual Travel Protect Platinum plan has low number of Claims registered whereas commission drawn is very high

EDA - DISTRIBUTION OF CLAIM AND COMMISSION AS PER AGENCY



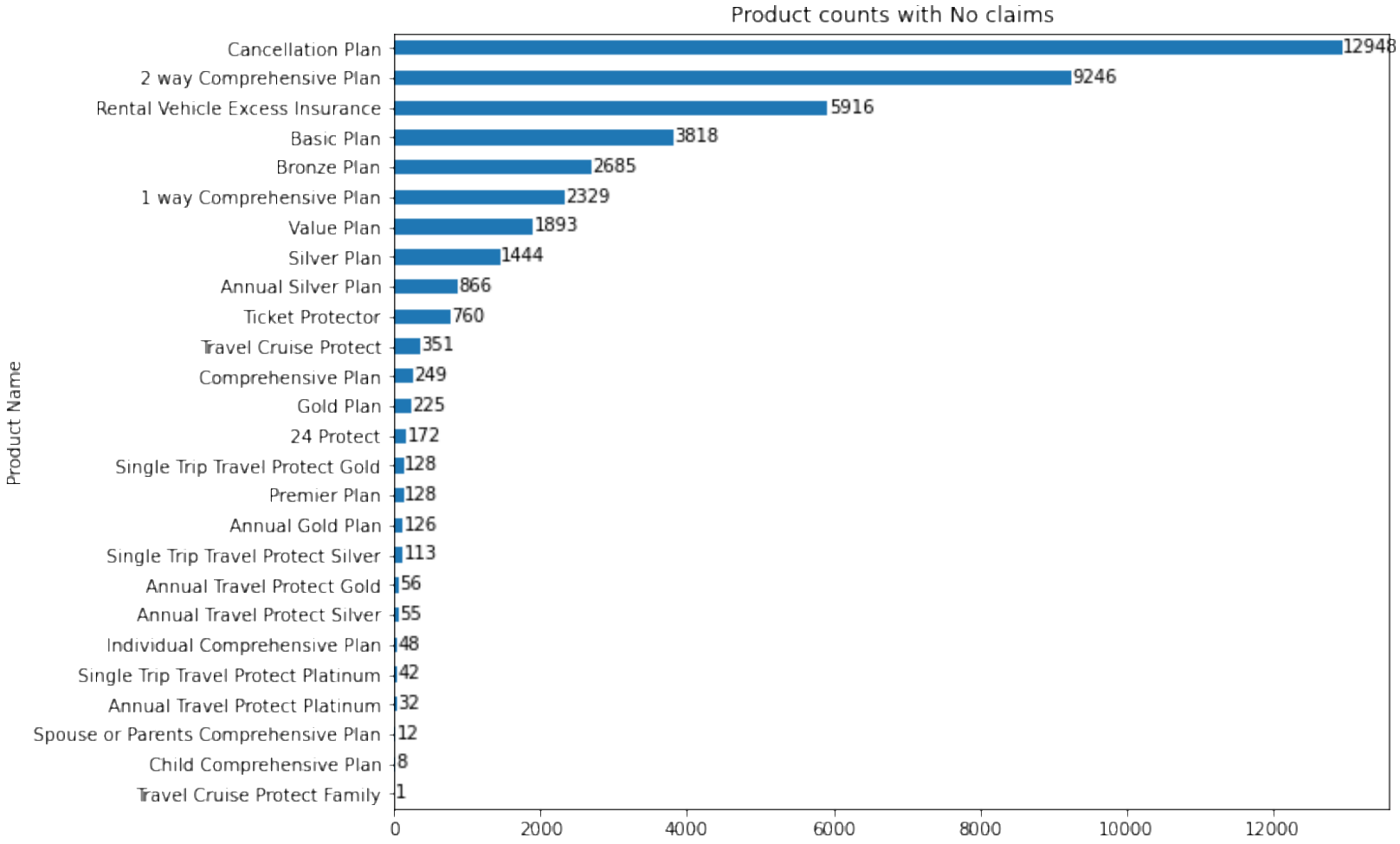
- Certain agencies have less registered claims but they draw high commission on the Insurance policies (CWT, JWT, JZI, RAB, TST)
- C2B & TTW have higher percent of registered claims but the mean commission is very low
- LWC have higher claim rate & mean commission

EDA - RELATION BETWEEN REGISTERED CLAIMS AND PRODUCT NAME



Bronze Plan, Annual Silver Plan, and Silver Plan has highest claims reported.

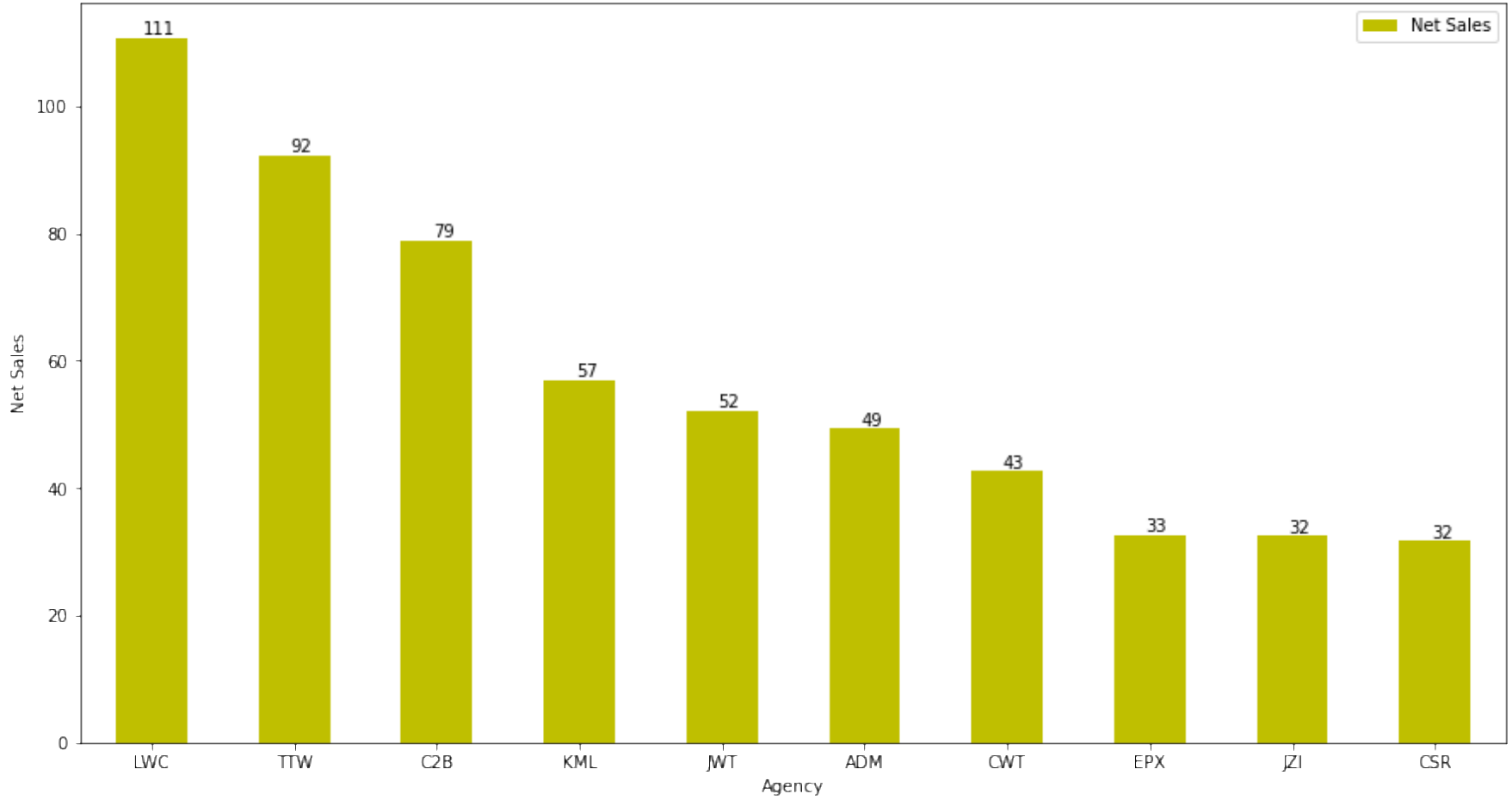
EDA - PRODUCT VS NET SALES



- As Cancellation Plan and 2 way comprehensive plans are contributing more in Sales.

- Even Cancellation plan has less claims so Agencies should focus more on this plan.

EDA - AGENCIES WITH MAX NUMBER OF NET SALES



The Agencies with Top 3 Net Sales were LWC, TTW and C2B.

MODELS AND APPROACHES

Model - Claim Detection

- Class 0 - No Claim -> Negative
- Class 1 - Claim -> Positive

Model akan berfokus di Recall

Model yang coba diantaranya

- Logistic Regression
- KNN
- Decision Tree Classifier
- Random Forest Classifier
- Support Vector Classifier

Diantara model tersebut, Decision Tree Classifier terpilih untuk di proses ke tahap selanjutnya karena score recall, precision, dan F1 lebih promising dibanding model lain

DT juga mampu mengenali kelas minoritas dan mayoritas

GOALS : Meminimalkan False-Negative (prediksi No Claim tetapi aktual Claim)

MODEL SCORE (VANILLA MODELS)

	Accuracy	Recall	Precision	F1
Logistic Regression TrainingCV (Mean)	0.984457	0.000000	0.000000	0.000000
Logistic Regression TrainingCV (Std)	0.000180	0.000000	0.000000	0.000000
Logistic Regression TestingCV (Mean)	0.984601	0.000000	0.000000	0.000000
Logistic Regression TestingCV (Std)	0.000004	0.000000	0.000000	0.000000
KNN TrainingCV (Mean)	0.984001	0.000000	0.000000	0.000000
KNN TrainingCV (Std)	0.000189	0.000000	0.000000	0.000000
KNN TestingCV (Mean)	0.984715	0.007407	0.200000	0.014286
KNN TestingCV (Std)	0.000226	0.014815	0.400000	0.028571
DT TrainingCV (Mean)	0.968459	0.047944	0.043396	0.047410
DT TrainingCV (Std)	0.001639	0.028098	0.027189	0.021680
DT TestingCV (Mean)	0.968632	0.007407	0.008333	0.021757
DT TestingCV (Std)	0.002549	0.014815	0.016667	0.017839
RF TrainingCV (Mean)	0.982461	0.012912	0.090918	0.025666
RF TrainingCV (Std)	0.000967	0.011043	0.058135	0.012851
RF TestingCV (Mean)	0.982548	0.000000	0.000000	0.012500
RF TestingCV (Std)	0.000454	0.000000	0.000000	0.025000
SVC TrainingCV (Mean)	0.984543	0.000000	0.000000	0.000000
SVC TrainingCV (Std)	0.000070	0.000000	0.000000	0.000000
SVC TestingCV (Mean)	0.984601	0.000000	0.000000	0.000000
SVC TestingCV (Std)	0.000004	0.000000	0.000000	0.000000

MODEL TUNING (RANDOMIZED SEARCH CV)

Benchmark

	precision	recall	f1-score	support
0	1.00	1.00	1.00	34523
1	1.00	0.89	0.94	542
accuracy			1.00	35065
macro avg	1.00	0.95	0.97	35065
weighted avg	1.00	1.00	1.00	35065

	precision	recall	f1-score	support
0	0.99	0.98	0.98	8632
1	0.06	0.07	0.06	135
accuracy			0.97	8767
macro avg	0.52	0.53	0.52	8767
weighted avg	0.97	0.97	0.97	8767

After Tuning: RSCV(DT with oversampling)

	precision	recall	f1-score	support
0	0.84	0.76	0.79	43155
1	0.78	0.85	0.81	43155
accuracy			0.80	86310
macro avg	0.81	0.80	0.80	86310
weighted avg	0.81	0.80	0.80	86310

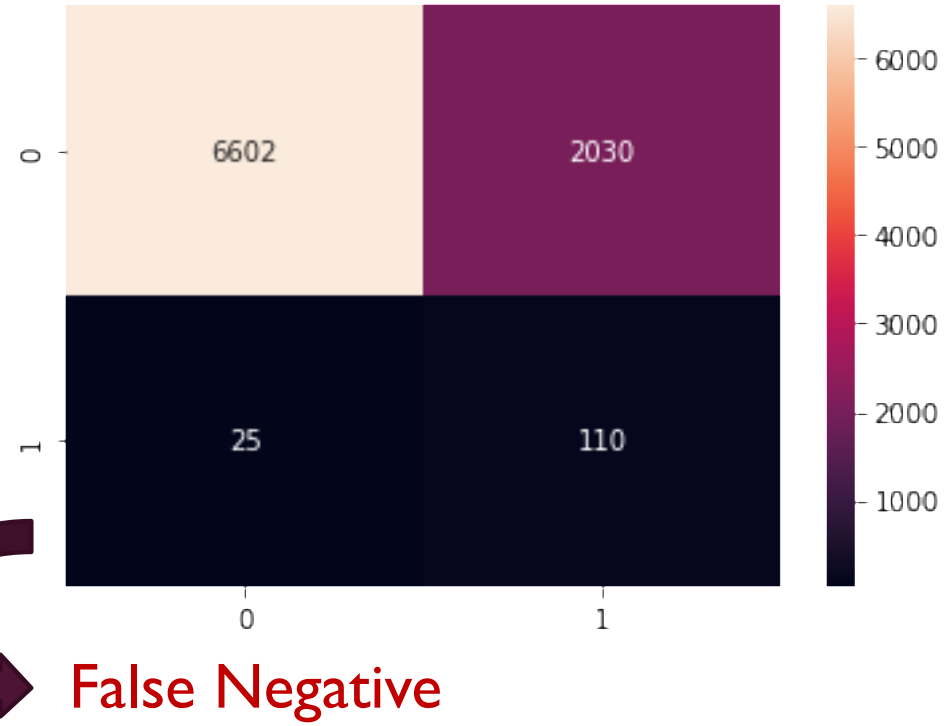
	precision	recall	f1-score	support
0	1.00	0.76	0.87	8632
1	0.05	0.81	0.10	135
accuracy			0.77	8767
macro avg	0.52	0.79	0.48	8767
weighted avg	0.98	0.77	0.85	8767

CONFUSION MATRIX

Benchmark

False Negative

After Tuning



False Negative

KESIMPULAN & REKOMENDASI

Kesimpulan

- Berdasarkan data, Cancellation Plan adalah *most selling* produk dengan tingkat claim rendah
- Berdasarkan data, Bronze Plan adalah produk dengan claim rate tertinggi, diikuti Annual Silver Plan, dan Silver Plan dibanding insurance plan yang lain
- Hypertuning RandomizedSearchCV terhadap model Decision Tree Classifier mampu memperkecil FN dibandingkan Benchmark

Rekomendasi

- Memfokuskan penjualan asuransi pada produk Cancellation Plan(guna mensiasati kenaikan revenue perusahaan karena sedikit yang claim)
- Meningkatkan premi produk agar cash flow lebih besar (guna subsidi tingkat claim yang cukup tinggi)
- Menambahkan fitur lain agar dapat lebih optimal (contoh terhadap netsales yang memiliki nilai negative) serta mencoba algoritma ML lain dan alternative pendekatan bisnis dengan menggunakan matrix lain

END OF SLIDE

- Thank you