Greetings!

My name is Fakhri Nurrahmadi,

A Data Analyst / Data Scientist with roughly 1 year of experience, most recently in the Finance industry, with a Data Science Certification.

Actively looking for new challenges to continue building towards a solid data skillset ☺

Link to Website

FAKHRI NURRAHMADI

DATA ANALYST

Hard Skills:

- Excel / Sheets
- Python
- SQL
- Tableau / Data Studio

Topics:

- Data Analysis
- Business Intelligence
- Data Science

JAKARTA POPULATION DATA

The following dashboard was created in attempt to display the continuous increase of Jakarta's population. The data is obtained from the Jakarta Open data website.

The data only shows up to the number of family registries, however we can still see that an upward trend of family registries per year in Jakarta.

The dashboard allows you to filter the data by the year, the city, and/or the district.

Tools used:
- Json files
- Python
- Google Sheets
- Google Data Studio

Link to Documentation

Link to Dashboard

Jakarta Open Data

AVERAGE KK PER TAHUN
3,291,408

Tahun | Kabupaten / Kota | Kecamatan

KK per Kecamatan

Kecamatan	# Kelurahan	# RW	# RT
Cakung	7	87	991
Cempaka Putih	3	30	373
Cengkareng	6	85	1,034
Cilandak	5	46	475
Cilincing	7	86	1,017
Cipayung	8	56	505
Ciracas	5	49	604
Duren Sawit	7	95	1,104
Gambir	6	44	482

KK per Tahun

Jumlah KK: 3.1M (2011), 3.2M (2012), 3.1M (2013), 3.1M (2014), 3.3M (2015), 3.3M (2016), 3.3M (2017), 3.5M (2018), 3.6M (2019)

Cohort Analysis - Sales

Sum of Price	month														Grand Total	Average
Years	CohortGroup	0	1	2	3	4	5	6	7	8	9	10	11	12		
⊟2018	Dec	573,310	145,096	162,643	166,756	182,446	156,598	178,542	167,995	168,744	188,811	204,500	283,072	176,973	2,755,486	211,960
⊟2019	Jan	289,667	47,299	37,589	47,321	53,389	56,921	41,784	54,681	63,327	83,821	80,364	24,415		880,578	73,382
2019	Feb	192,442	24,555	48,399	40,566	39,293	28,850	39,505	48,781	44,094	55,905	8,605			570,994	51,909
2019	Mar	227,497	28,924	40,943	37,411	51,746	40,212	38,716	52,545	70,056	26,869				614,920	61,492
2019	Apr	135,581	28,562	18,991	12,233	23,918	36,266	31,280	46,102	9,455					342,388	38,043
2019	May	133,874	14,041	32,692	12,246	41,399	35,055	35,719	18,546						323,573	40,447
2019	Jun	132,641	16,009	13,498	24,022	26,182	36,611	7,298							256,261	36,609
2019	Jul	136,150	7,580	18,682	22,036	30,254	7,729								222,431	37,072
2019	Aug	90,918	30,564	38,736	28,071	1,326									189,615	37,923
2019	Sep	140,028	39,791	41,397	5,025										226,241	56,560
2019	Oct	174,830	35,504	7,575											217,908	72,636
2019	Nov	165,833	8,217												174,050	87,025
2019	Dec	16,916													16,916	16,916
Grand Total		2,409,686	426,141	461,145	395,688	449,953	398,243	372,844	388,651	355,676	355,405	293,469	307,487	176,973	6,791,361	522,412

COHORT ANALYSIS - SALES

A cohort analysis is an analytical technique which involves categorizing groups of people and monitoring their changes over time.

In this particular case, a cohort analysis is performed to see the sum of sales amount of each cohort group over time.

The analysis can be found in the documentation

Tools used:
- Csv file
- Microsoft Excel

Link to Documentation

Link to Dashboard

CUSTOMER CHURN PREDICTION

By way of machine learning modelling, we can predict customers that have tendencies to churn based on their data.

In this particular case study, customer churn prediction is performed for a telecommunication company to retain valuable customers.

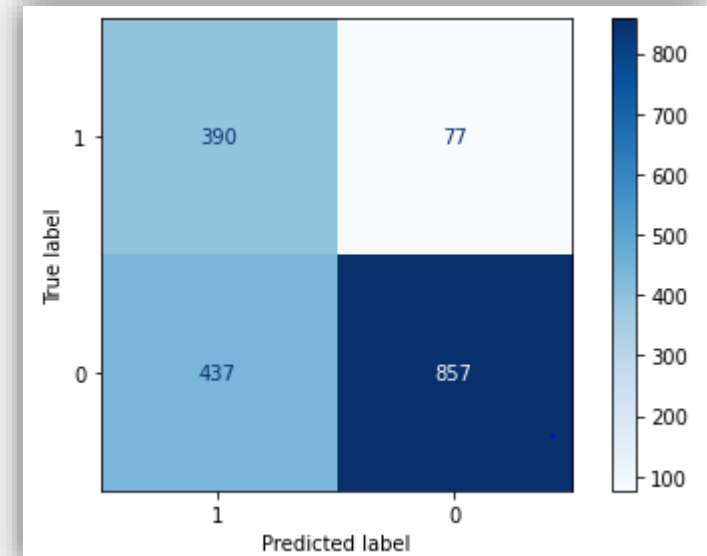A discount program is devised to retain churning customers.

Discounts are only given to churning customers that are valuable, and their value is calculated using regression analysis.
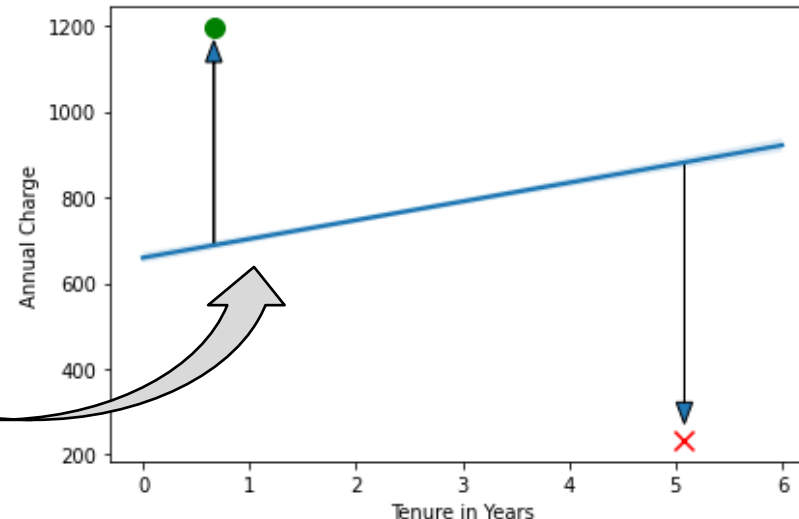
Tools used:
- Python

Classification Report

Confusion Matrix

A Gaussian Naïve Bayes model was used to predict churning customers.

The model performs with a recall score of 84% for churning customers, meaning that there is only a 16% chance of False Negatives. Recall score is used as the metrics since the concern is to ensure as many churning customers are retained.

ELT - ADVENTUREWORKSDW

| Data source | → | Data Lake | → | BI |

Tools used:
- Microsoft SQL Server as Data Source
- PostgreSQL on Google Cloud SQL as Data Lake
- Python
- DBeaver

This project was done in attempt to simulate the ELT process from a single data source to a data lake. In this particular case, raw data was extracted from one database and loaded to the data lake where it can then be transformed to get the reports necessary.

Future works:
- Implement Airflow to automate the process of extract and load
- Create the BI dashboard

Link to Documentation