

SQL problems and Cross-sell/Upsell Future Forecast

Solutions

By : Khalisul Akbar

Problem 1 : TOP 3 PROVINCES WITH THE MOST BAD CUSTOMERS

```
# Problem 1 : Top 3 Province with the mose bad customer
SELECT provinces_name, COUNT(is_bad_cust) AS total_bad_cust
FROM Astra_data
GROUP BY provinces_name
ORDER BY total_bad_cust DESC
LIMIT 3;
```

Result :

Result Grid			Filter Rows:
	provinces_name	total_bad_cust	
▶	JAWA BARAT	6029	
	JAWA TIMUR	3232	
	DKI JAKARTA	2758	

Problem 2a: UPSELL

```
# Problem 2 : Top 3 provinces with the most Upsells
SELECT provinces_name,
CASE
    WHEN first_bu = second_bu AND (first_bu AND second_bu) is
NOT NULL THEN 'Up Sell'
    ELSE 'Cross Sell'
END AS strategy,
COUNT(CASE
    WHEN first_bu = second_bu AND (first_bu AND second_bu) is
NOT NULL THEN 'Up Sell'
    ELSE 'Cross Sell'
END) AS total
FROM mytable
WHERE first_bu = second_bu
GROUP BY provinces_name
ORDER BY total DESC
LIMIT 3;
```

Result :

Result Grid				Filter Rows:
	provinces_name	strategy	total	
▶	JAWA BARAT	Up Sell	2702	
	JAWA TIMUR	Up Sell	1463	
	DKI JAKARTA	Up Sell	1392	

Problem 2b: CROSS-SELL

```
# Problem 2 : Top 3 provinces with the most Cross-sells
SELECT provinces_name,
CASE
    WHEN first_bu = second_bu AND (first_bu AND second_bu) is
NOT NULL THEN 'Up Sell'
    ELSE 'Cross Sell'
END AS strategy,
COUNT(CASE
    WHEN first_bu = second_bu AND (first_bu AND second_bu) is
NOT NULL THEN 'Up Sell'
    ELSE 'Cross Sell'
END) AS total
FROM mytable
WHERE first_bu != second_bu
GROUP BY provinces_name
ORDER BY total DESC
LIMIT 3;
```

Result :

Result Grid				Filter Rows:
	provinces_name	strategy	total	
▶	JAWA BARAT	Cross Sell	3327	
	JAWA TIMUR	Cross Sell	1769	
	DKI JAKARTA	Cross Sell	1366	

Problem 3: MONTHLY GTV

```
# Problem 3 : Monthly Gross Transaction Value (GTV)
SELECT year_, month_, SUM(asset_price ) as GTV , 'Price in
million (x10^6)' AS unit
FROM (
    SELECT YEAR(a.first_go_live_date) AS year_
    ,MONTH(a.first_go_live_date) AS month_, a.first_asset_price AS
asset_price
    FROM mytable AS a
    UNION ALL
    SELECT YEAR(b.second_go_live_date) AS year_,
    MONTH(b.second_go_live_date) AS month_, b.second_asset_price AS
asset_price
    FROM mytable AS b) AS c
GROUP BY year_, month_
ORDER BY year_, month_
```

Result :

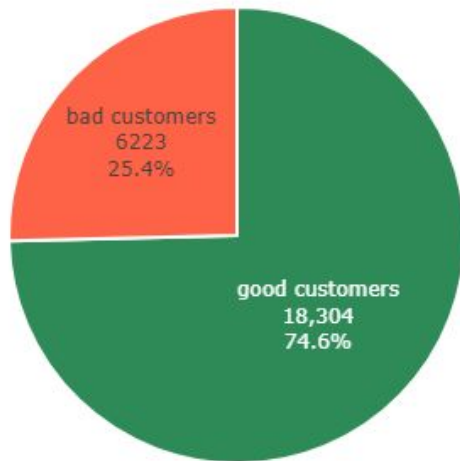
	year_	month_	GTV	unit
▶	2018	1	181984	Price in million (x10^6)
	2018	2	131686	Price in million (x10^6)
	2018	3	158497	Price in million (x10^6)
	2018	4	198198	Price in million (x10^6)
	2018	5	200749	Price in million (x10^6)
	2018	6	154398	Price in million (x10^6)

Cross-sell and Upsell Forecast with Random Forest Regressor

Exploratory Data Analysis (EDA)

Bad Customers

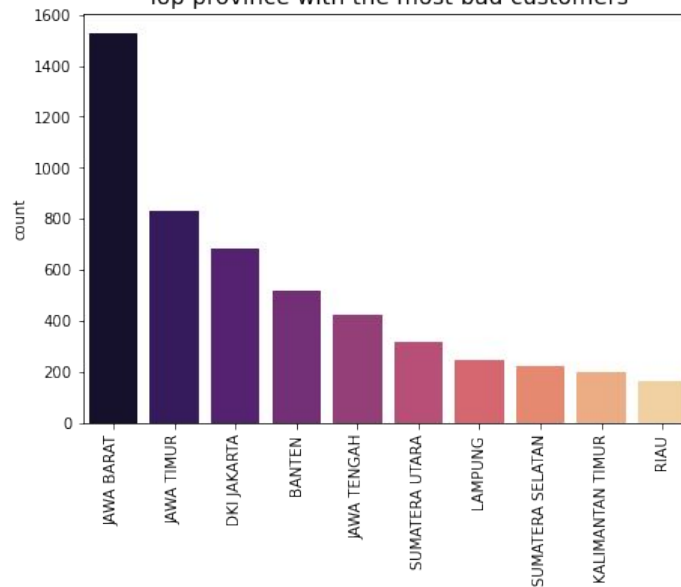
Good customers vs Bad customers



Bad Customers Proportion

Overall, there are approximately one-quarter of total customers are bad customer.

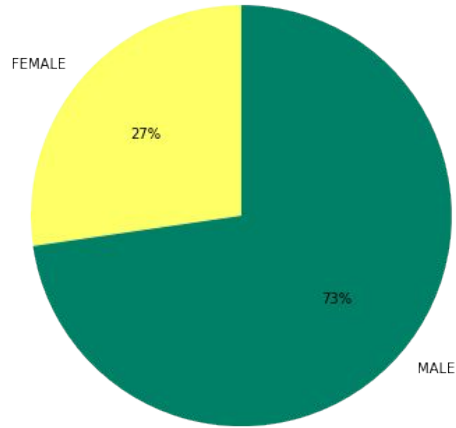
Top province with the most bad customers



Most of them come from **Jawa Barat**.

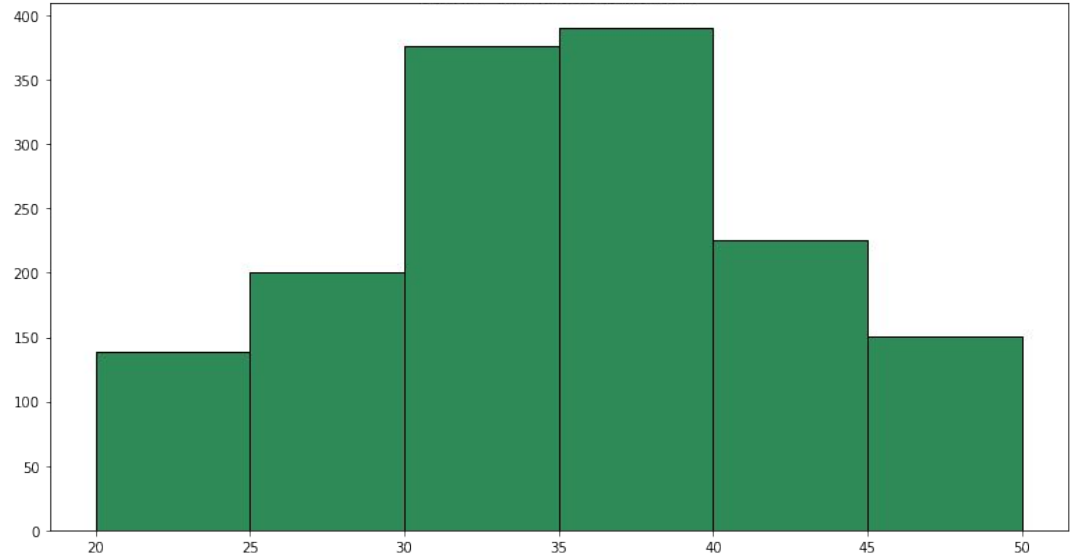
Bad Customers Age and Gender

Bad customers gender proportions



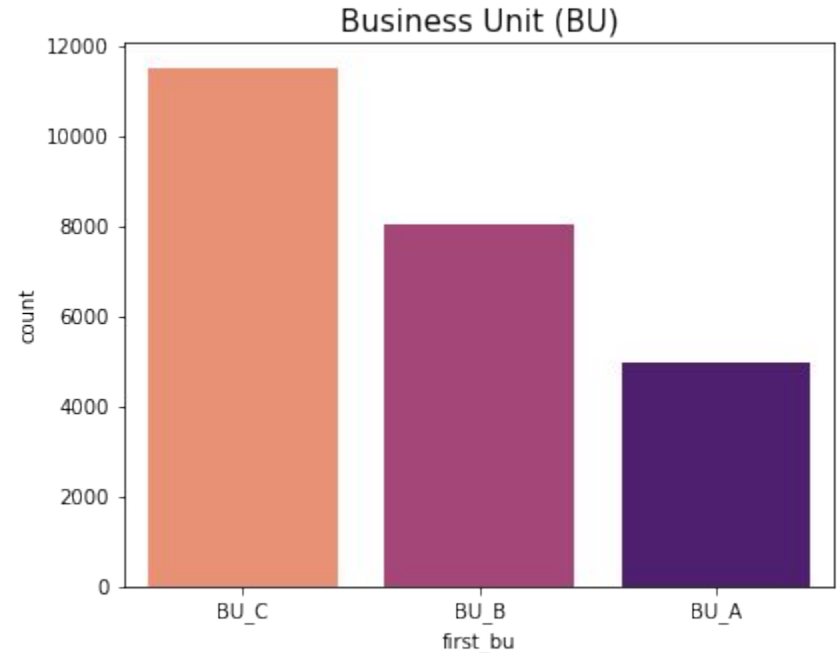
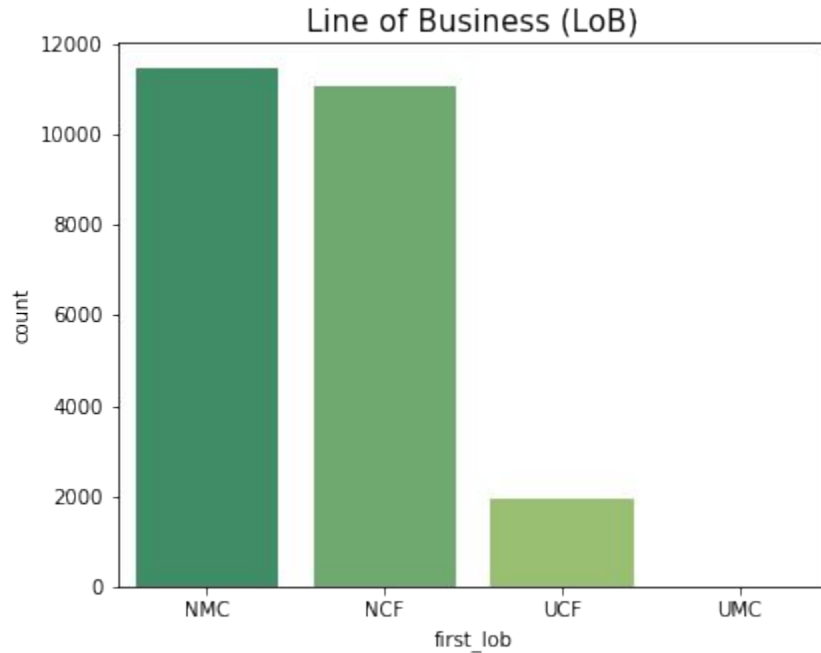
Male customers are likely to become bad customers

bad customers average age



Age range : **20 to 50 years old**
Average Age : **30 to 40 years old**

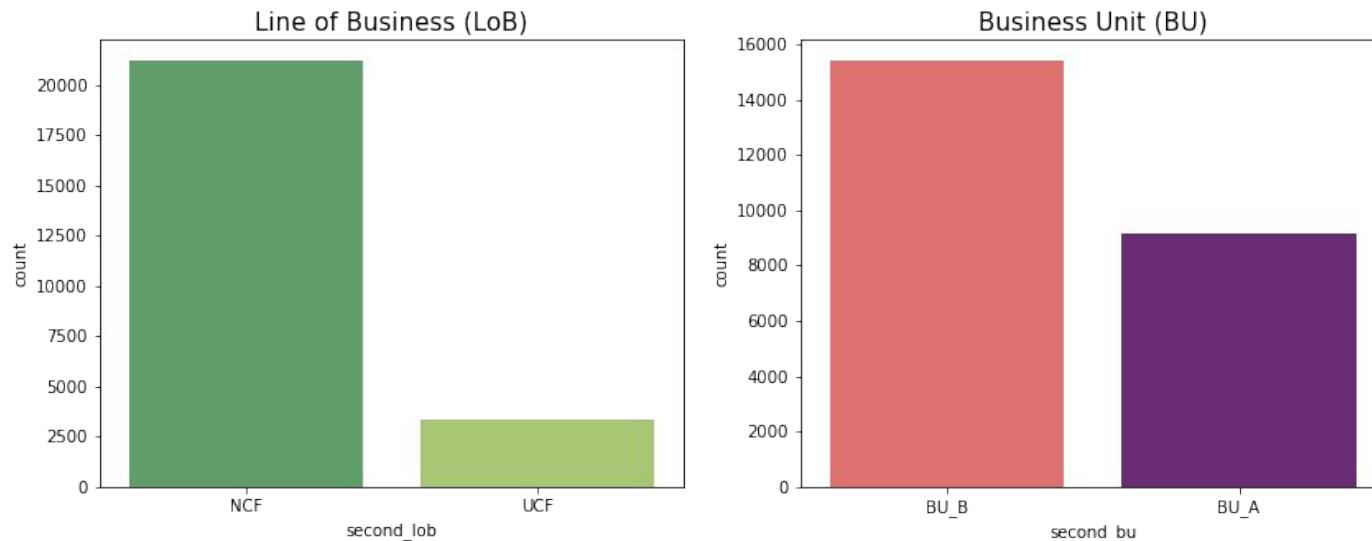
Popular Line of Business and Business Unit



The most popular Line of Business chosen as the customers first transaction are **NMC (New Motor Cycle)** and **NCF (New Car Finance)**

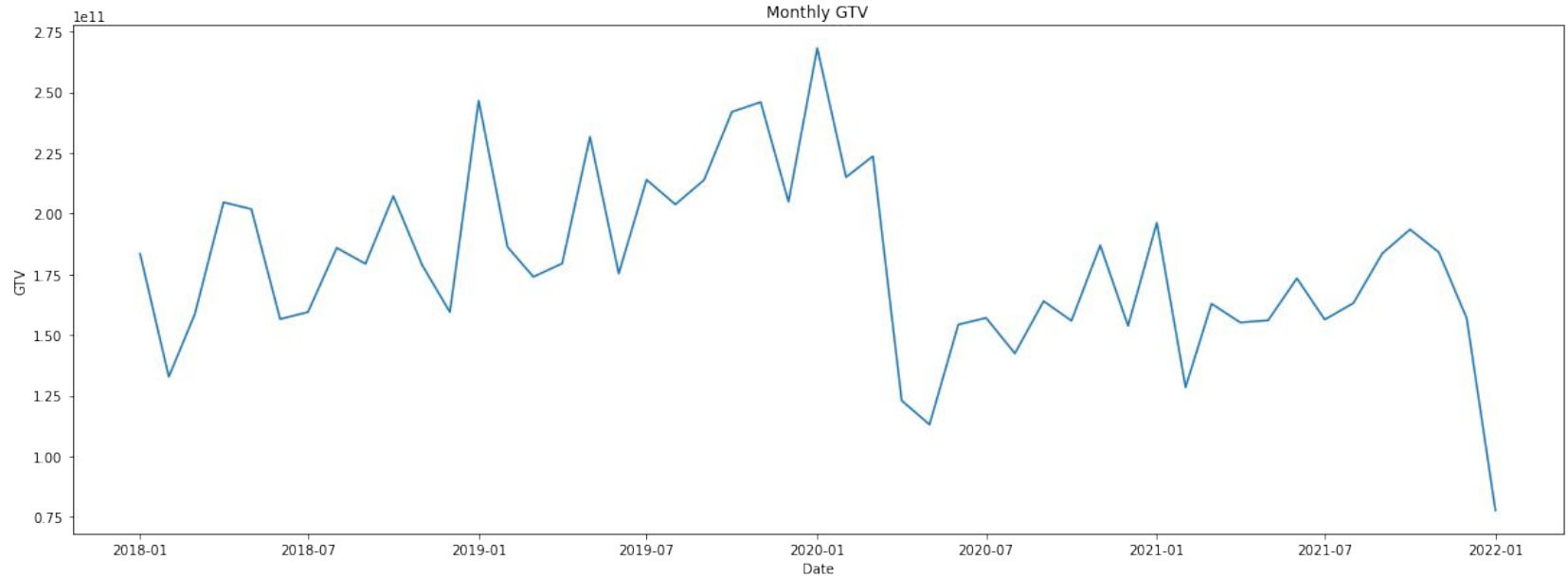
In their first transaction, most customers prefer to go to **BU_C** business unit

and...



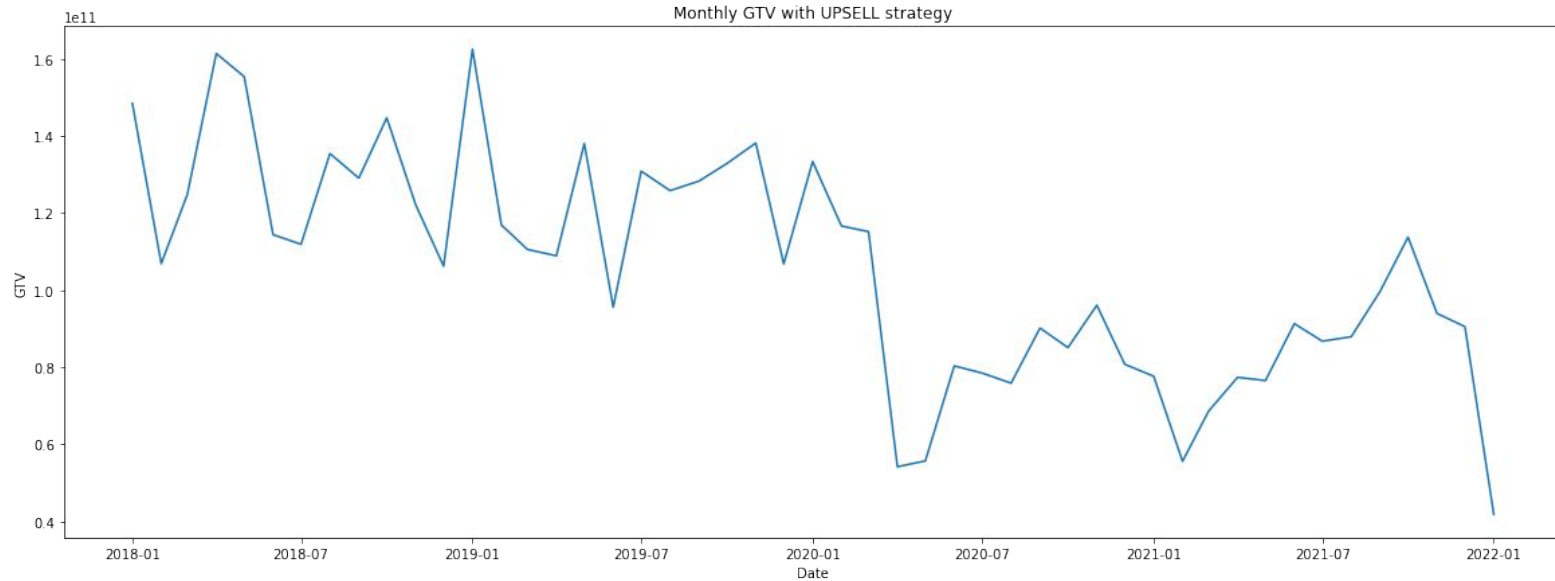
As for their second transaction, they prefer **NCF (New Car Finance)** and most likely will do the transaction at **BU_B business unit**.

Monthly Gross Transaction Value (GTV)



The GTV is in a **increasing trend** at the start of 2018 to the early of 2021. However, it has been **declining since the start of 2020**. Accidentally, early 2020 was when the global pandemic hit. There might be some correlation between the effect of global pandemic and the company sales.

Monthly GTV for UPSELL Strategy



Has been declining from the start of 2018.

Monthly UPSELL Strategy Forecast

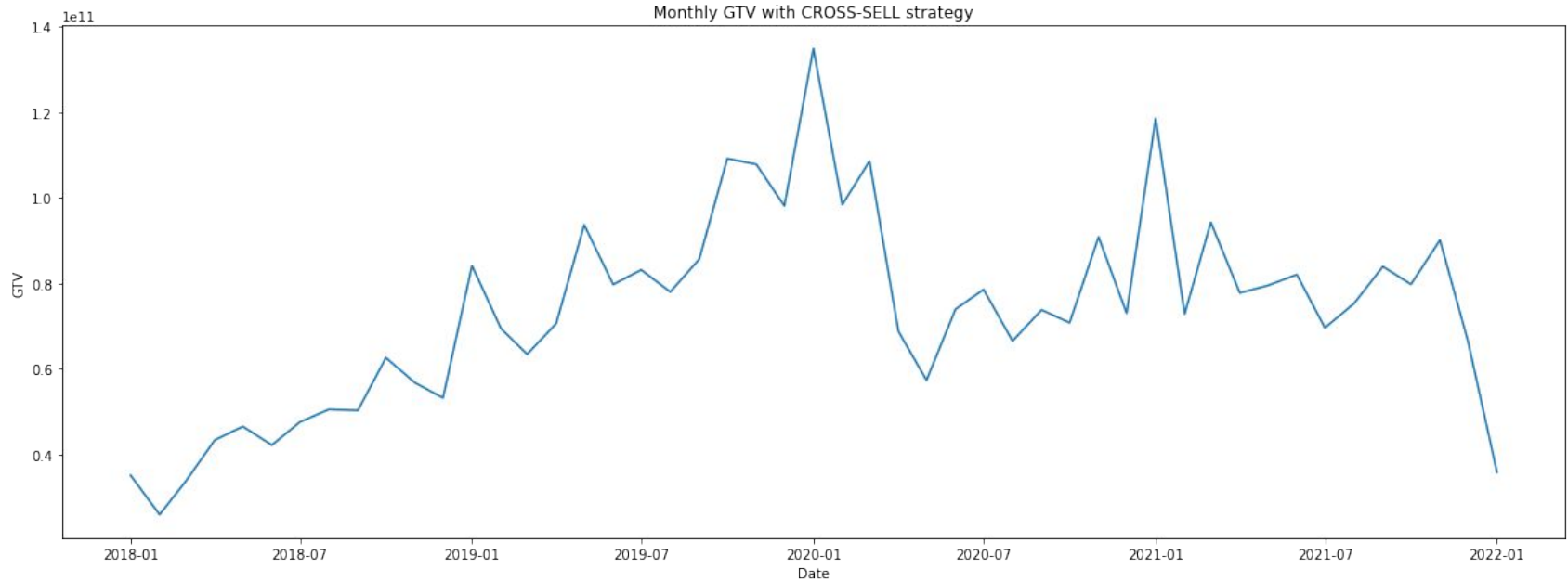
Next Month Upsell GTV forecast Price = IDR 85687250000.00, RMSE : IDR 25953618702.27



- The latest month from the dataset is : 01 january 2022, and the gtv generated from **UPSELL** strategy is : **41.86 Billion**
- next month is : 01 feruary 2022, and **the GTV forecasted** by the model is **85.68 Billion**

the model forecasted future GTV will be INCREASING

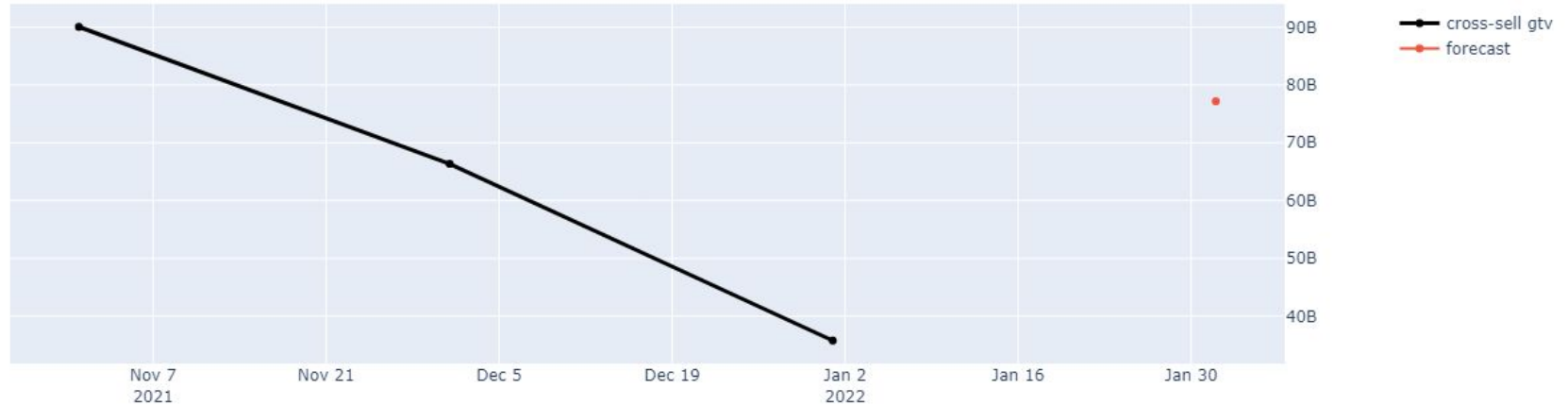
Monthly GTV for CROSS-SELL Strategy



The GTV generated from this strategy was **increasing in the beginning** and even reached its **max GTV** value of around **130 Billion Rupiah**. Since then the GTV has been declining and generated almost similar GTV as early 2018.

Monthly CROSS-SELL Strategy Forecast

Next Month Upsell GTV forecast Price = IDR 77189160000.00, RMSE : IDR 26026687343.03



- The latest month from the dataset is : 01 january 2022, and the gtv generated from **CROSS-SELL** strategy is : **35.77 Billion**
- next month is : 01 feruary 2022, and the GTV **forecasted** by the model is **77.189 Billion**

the model forecasted future GTV will be INCREASING

Conclusion :

- From the total customers there are 18,304 Good customers and 6,233 Bad Customers. The total percentages of **Bad Customers are 25.4 %**.
- From 6,233 Bad Customers, Most of them come from :
 1. **Jawa Barat (1,528)**
 2. **Jawa Timur (829)**
 3. **DKI Jakarta (686)**
- Overall, the characteristic of the bad customers are :
 1. **Male customers** are the most likely to become **bad customer** (customer ever overdue 60 days/ ever repo/ ever WriteOff)
 2. **Average age range** of the bad customers are **34 to 35 years old**.
 3. They are likely to be from **kabupaten Bekasi** and **province Jawa Barat**.
 4. They are likely to buy from **BU_B** dan **BU_C**.
 5. They are likely to purchase from the Line of Business unit : **New Motor Cycle(NMC)** and **New Car Finance(NCF)**

- **Most Popular :**

1. Most customers prefer line of business (lob) NMC and Business Unit (BU) BU_C as their **first transaction priority**.
2. The most preferred lob and BU for theirs **second transaction** are : NCF for the lob and BU_B for the BU.
3. NCF as the first_lob and NCF as the second_lob are the most popular combination of lob **being purchased by customers at the same time**.

- **Overall Monthly GTV is in the downtrend.** Although the GTV generated monthly from the start of 2018 up until January 2020 showed an increased trend. However, if we take a closer look at current GTV, it is significantly lower than the start of the data or January 2018. To conclude, there might be something happening in the duration of early 2020 to 2022 which caused the GTV to be lower and lower (possibly related to COVID-19) which caused the overall sales to be decreasing.

- **However, if we are looking by its strategy :**

UPSELL : we can see that the overall trend for monthly transactional value using Upsell strategy is in Downtrend. It has been declining from the beginning of 2018.

CROSS-SELL : the GTV has been increasing since the beginning of the 2018 until the start of 2020 . However, since the beginning Of 2021 - 2022 the GTV generated using this strategy is has been decreasing.

- **Model Peformance and Future Improvement**

1. The model is trained using RANDOM FOREST REGRESSOR
2. The model performance need to be improved because the RMSE score shown by UPSELL Model and CROSS-SELL Model is still very high.
3. Future improvement is going to be :
 - Hyperparameter Tuning to look for the best Hyperparameter for each model.
 - Looking for another better Algorithms that is more suitable to solve time-series problem such as : ARIMA, PROPHET, LSTM, etc

Thank You

