



DATA SCIENCE ACADEMY CAPSTONE PROJECT RIDE HAILING INTERNET PACKAGE

GROUP 12

Yustinus Kunta Wibisana Shelby Marsa Istiqomah Wahyu Sejati Roso Rizaldy Al Kautsar Utomo



Executive Summary of DSA -Ride Hailing Internet Package for Driver



- The growth of online ride-hailing business in Indonesia is potential market to get more data users for Telkomsel
- Telkomsel launched a special combo data package for online ride-hailing drivers with IDR 75K for price
- Competitor offered merely same package for "mitra driver" with competitive price (IDR 20K 75K)



- 1 Only 41.77% (547 K subscriber) takers from 1.2 Mio mitra driver who use Telkomsel number
- 2 How to identify takers from potential mitra driver?
- How much affordable price for combo package for mitra driver?

Objective

- 1 Build supervised model that can predict takers
- 2 Create segment of customer with clustering
- 3 Develop **subsegment** based on ARPU



Proposed Solutions

- 1 Identify potential takers with **random forest** classification algorithm
- 2 Clustering:
 - 1. **Economic Customer** = Lowest ARPU & data consumption, Usage primarily for Ojol apps.
 - 2. Mid Spender = Medium ARPU & payload, highest voice mou & transaction
 - **3.** Data Addict = High data consumption & dominant video & social app usage.
- 3 Create package based on selected sub-segment



Result

- There are potential 55k new numbers of takers based on this algorithm. In total there are 406k potential package takers. 44.2% taker rate, uplift +4% from previous data.
 - AUC 93.8%, F1-Score 91% with random forest
- Develop 3 main cluster with 0.63 silhouette score, and distinctive usage of data, voice, and sms
 - The cluster are Economic Customer, Mid Spender, and Data Addict
- Created 4 package out of 9 sub-segment that can bring uplift of IDR 4937/subs

S Business Benefit

- 55,026 new takers (based on FP) can bring additional IDR 4.12
 Billions
- Median revenue of each cluster summed up into IDR 9.35 Billion from 78,917 subs
 - 4 New Package Offering: Ojol Hemat A, Ojol Hemat B, Ojol Gaspol, and Ojol Juara that can bring IDR 9.82 Billions from 78,917 subs. +67% of revenue compared to baseline IDR 75k package
 - Average uplift per subscriber reach IDR 4937

GROUP 12

Yustinus Kunta Wibisana Shelby Marsa Istiqomah Wahyu Sejati Roso Rizaldy Al Kautsar Utomo

CRISP-DM

RIDE-HAILING INTERNET PACKAGE

Business Understanding
Understand objective and requirement
from business perspective

Data Understanding
Getting familiar with the data to form hypotheses



Data Preparation
Construct dataset from raw data

Modelling
Building the model for desired output

EvaluationAssess the quality of the model based on requirement

Deployment

Put the result to work and achieve the goals







Ride-Hailing is **multi billion business** in Indonesia, and the driver stand as partner which plays crucial role, represent **0.9% of Telkomsel Population**

Active Ride-Hailing Drivers

1.48 Million

Serving

Driver to
Consumer
Ratio
1:11

Active Ride-Hailing Users

16.86 Million



34% of them have monthly income of > IDR 3.5 million after joined as ride hailing driver, only 8% of them already have it before join as a partner



Daily income of IDR 150-200k/day, communication expense can be a burden if Telco company don't provide the best offering



Telkomsel as connectivity provider stand as enabler to make the driver experience with ride-hailing app more seamless

XL

	Bulanan 1	Bulanan 2	Mingguan
Harga	Rp50.000	Rp75.000	Rp20.000
Masa Aktif	30 Hari	30 Hari	7 Hari
Kuota	11 GB	20 GB	2 GB
Gratis Aplikasi	Gojek/GoCar Driver & Waze	Gojek/GoCar Driver & Waze	Gojek/GoCar Driver & Waze
Kuota telepon ke sesama operator	Unlimited	Unlimited Unlimite	
SMS ke sesama operator	Unlimited	Unlimited	Unlimited
Kuota telepon ke operator lain	50 Menit	50 Menit	15 Menit
SMS ke operator lain	100 SMS	100 SMS	-

Price/GB

IDR 3750

Telkomsel

Tipe Layanan	Paket Swadaya Telkomsel		
Biaya per bulan	Rp75.000		
Masa periode aktif	30 Hari		
Kuota telepon ke sesama operator	Tidak terbatas atau unlimited		
Kuota telepon ke semua operator	200 menit		
SMS	500 SMS		
Kuota internet	15 GB		

Price/GB

IDR 5000

Indosat

Tipe Layanan	Paket Gaspol Swadaya Indosat		
Biaya per bulan	Rp50.000		
Masa periode aktif	30 hari		
Kuota internet	10 GB		
Telepon ke sesama Indosat	Gratis		
Telepon ke semua operator	Gratis 100 menit		

Price/GB

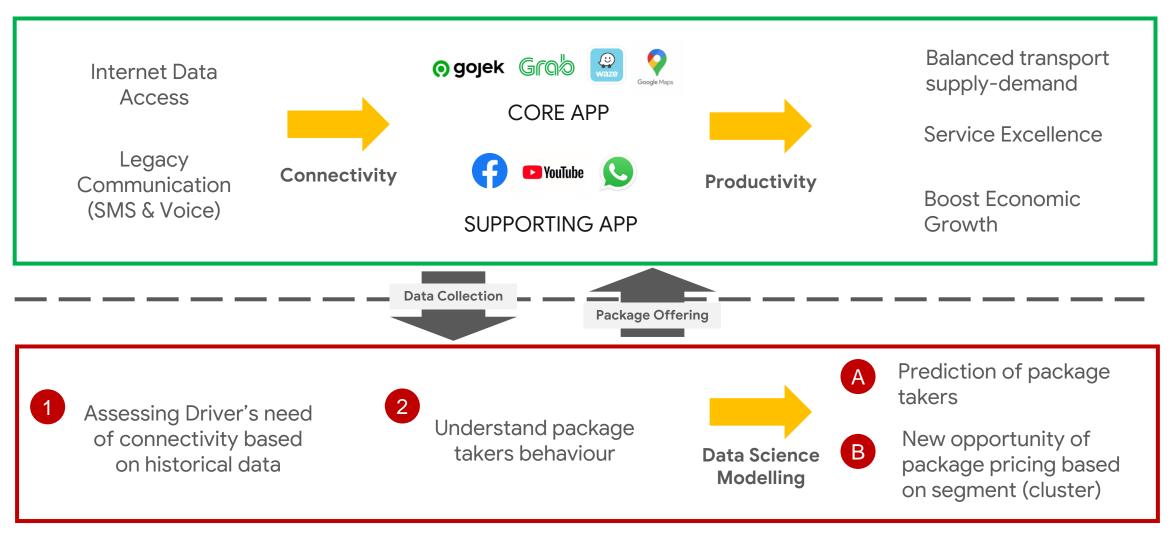
IDR 5000

Telkomsel has several competitor with **more competitive price for ride-hailing driver package.** We need to **enhance value proposition** to improve the takers of package



Understanding driver needs and usage of connectivity with data science can be the key to drive more takers in Ride-Hailing package

Driver Workflow



Telkomsel's Business Opportunity



... based on previous business problem, **3** objective and key result can be derived with data science process (classification & clustering)

Problem Statement

Telkomsel already have **ride-hailing package** for drivers, priced at IDR 75k/month.

The package are targeted for whitelisted MSISDN, as October'19 there are 1.31 mio of whitelist with 547k takers (41.7%)

Objective

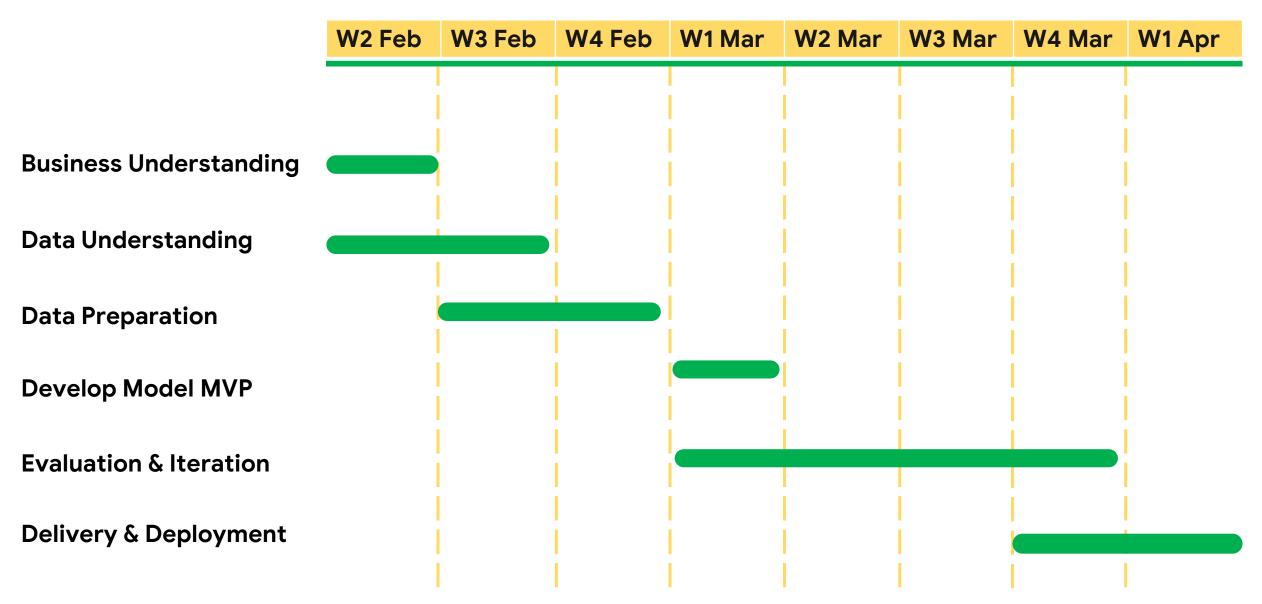
- 1 Build supervised model that can predict takers
- 2 Create segment of customer with clustering
- 3 Develop **subsegment** based on ARPU

Key Result

- 1 Achieve 80% Area Under the Curve (AUC) and 70% F1-Score
- 2 Reach Silhouette Score of >0.6 to determine cluster size
- Achieve IDR 3000 ARPU uplift/subs by offering subsegment



... to achieve the result, expected timeline is **2 months** of CRISP-DM complete cycle

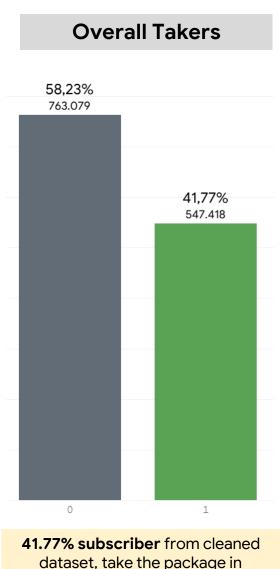


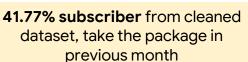


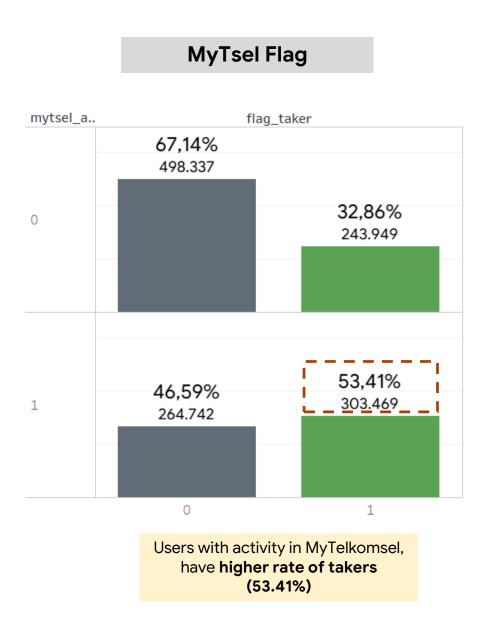




547k subscriber from whitelist took the package, and users with active MyTelkomsel app have higher rate of takers









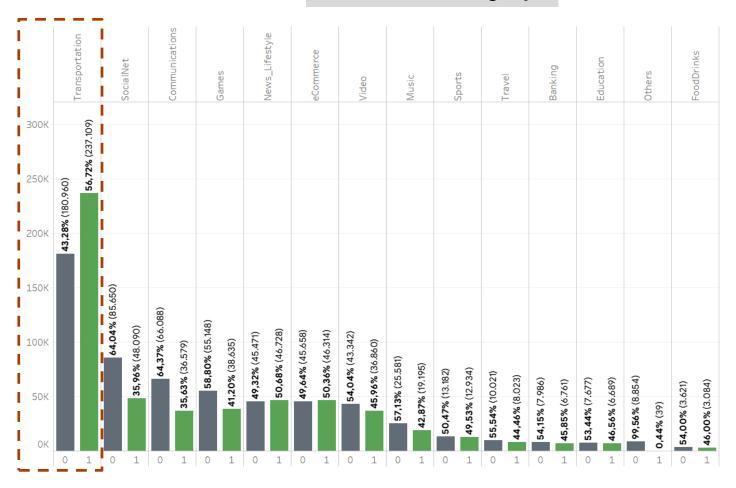
Region and most frequent app category of users also have distinguish rate of takers

Region

	flag_taker		
region =	0	1_	
05.Central Jabotabek	55,51%	44,49%	
05.Central Jabotabek	149.041	119.466	
06.Eastern Jabotabek	56,59%	43,41%	
00.Eastern Jabotabek	113.907	87.375	
09.Jatim	60,63%	39,37%	
09.Jaciiii	77.088	50.047	
04.Western Jabotab	55,50%	44,50%	
	62.758	50.318	
08.Jateng	59,68%	40,32%	
00.5ateng	66.788	45.123	
07.Jabar	58,42%	41,58%	
07.50001	64.582	45.971	
01.Sumbagut	62,51%	37,49%	
01.5dilibugut	53.797	32.259	
12.Sulawesi	55,78%	44,22%	
12.odiawesi	46.965	37.228	
03.Sumbagsel	61,75%	38,25%	
	40.749	25.240	
02.Sumbagteng	58,46%	41,54%	
	35.487	25.211	
11.Kalimantan	65,16%	34,84%	
	27.624	14.768	
10.Balinusra	65,15%	34,85%	
	19.819	10.600	
13.Puma	53,99%	46,01%	
	4.474	3.812	

Users from Jabotabek contribute to highest number of users and high takers rate (>43%)

First Rank Category



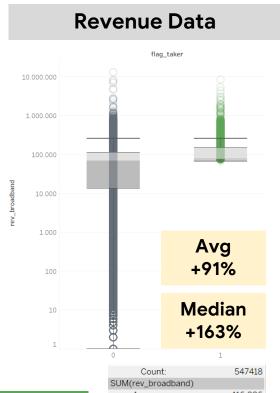
Users with high activity on **Transportation Apps have the highest rate of takers** with 56.72%



Taker

Non-Taker

Revenue Analysis – Package Taker have much higher revenue on the **data revenue** but fall short on **voice revenue** and **sms revenue**



Count:	547418
SUM(rev_broadband)	
Average:	116,096
Minimum:	69,116
Maximum:	8,499,000
Median:	77,400
First quartile:	75,000
Third quartile:	150,000
Count:	763079
SUM(rev_broadband)	
Average:	60,655
Minimum:	0
Maximum:	13,225,000

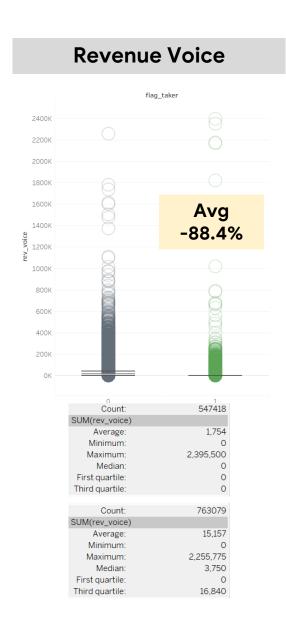
Median:

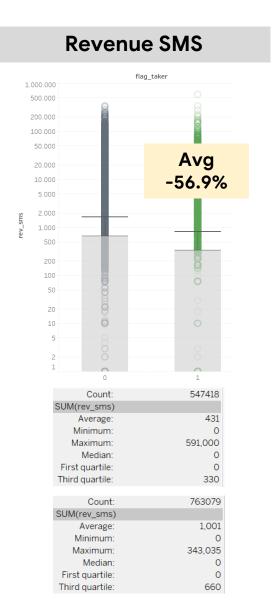
First quartile:

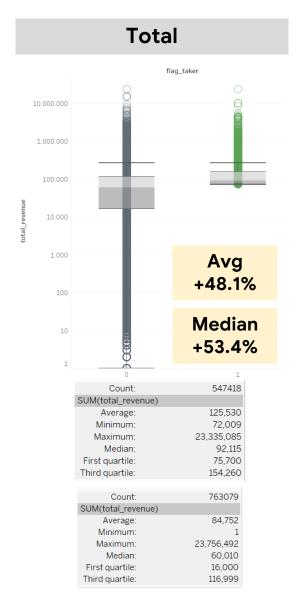
Third quartile:

29.320

91.028









Usage Analysis – Package Taker consistently have heavier usage on data payload, minutes of voice and number of sms



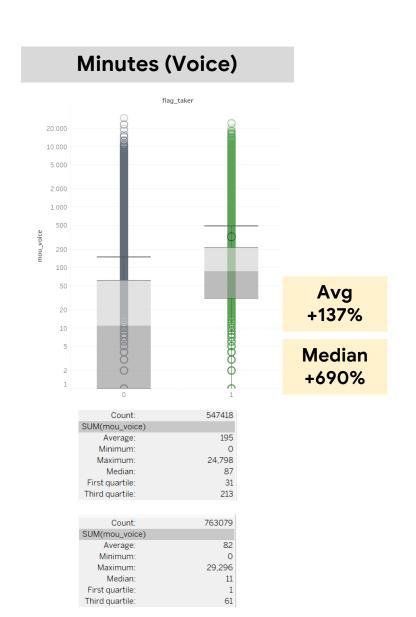
3,730,634

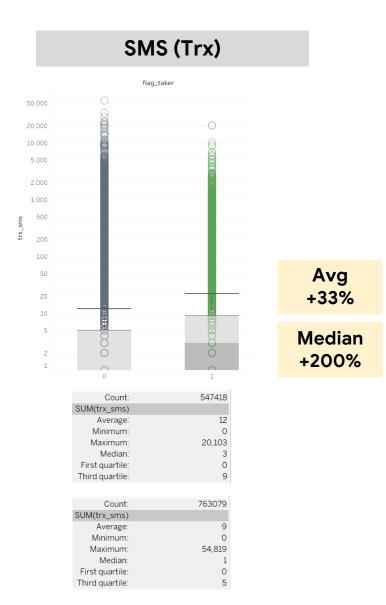
14,368,719

First quartile:

Third quartile

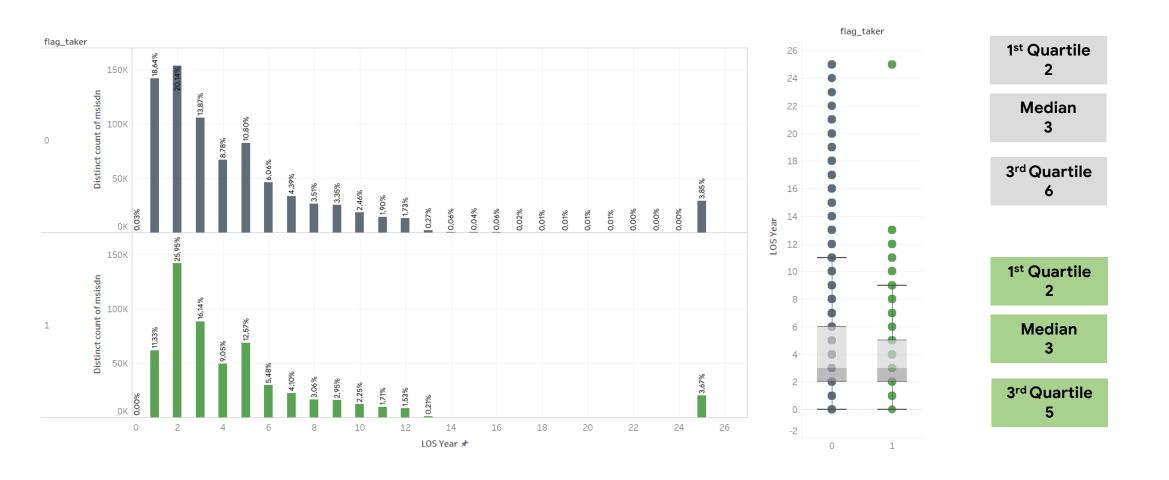
277







Length of Stay – Package Taker relatively have skew to younger stay in Telkomsel as the 3rd Quartile is 5 years

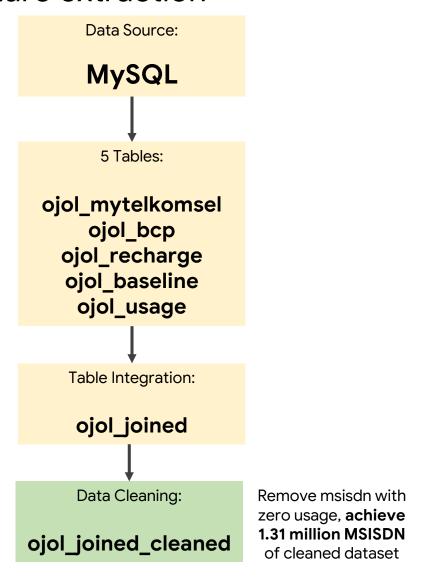


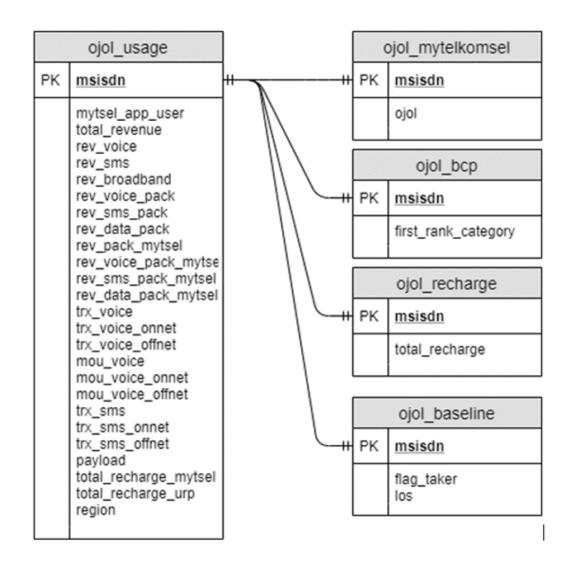






Data Pipeline and Integration – Encoding, Ratio, Flag, Binning are used for feature extraction





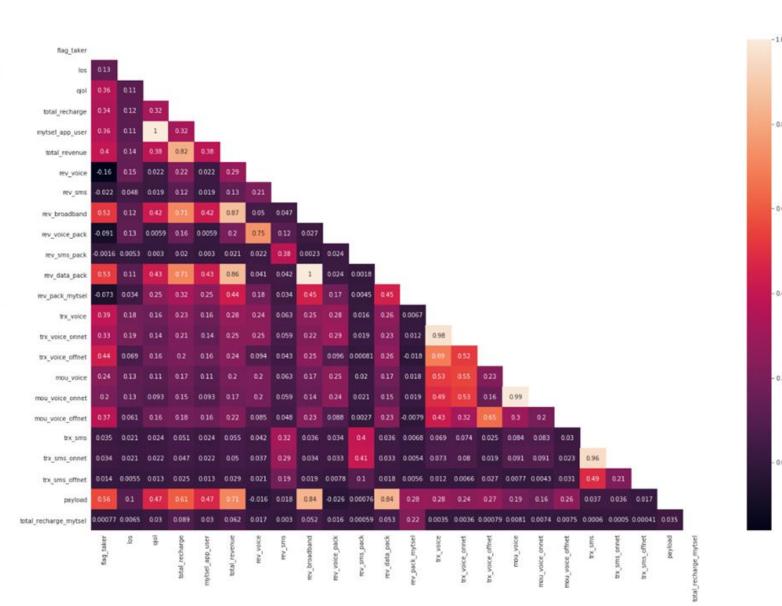


Understanding the Correlation

Top 10 Correlation

Feature	Abs Correlation
payload	0.559477
rev_data_pack	0.528849
rev_broadband	0.518248
trx_voice_offnet	0.436402
total_revenue	0.399883
trx_voice	0.390107
mou_voice_offnet	0.374851
ojol	0.358926
mytsel_app_user	0.358926
total_recharge	0.342274

Based on the exploration on correlation, variable from data usage such as payload, revenue data package, and revenue broadband have the highest absolute correlation to flag





Feature Extraction and Engineering – Encoding, Ratio, Flag, Binning are used for feature extraction

One Hot Encoding

One Hot Encoding on columns first_rank_category and region

Flag

Add flag whether customer have voice, sms, data transaction

Missing Value Handling

- Impute missing value on 'first_rank_category' with string 'Missing Value'
- Drop columns that only contain 0 values

Ratio

- Add log revenue due to large skewness on revenue data distribution
- Binning revenue on interval (250k,500k,1M,2.5M,5M,10M,15M,20M)

Binning and Log Value

- Add mou/trx ratio
- Add ratio for each service type revenue / total revenue
- Add ratio for each service type trx * los
- Add revenue/recharge ratio
- Add recharge revenue ratio

Total Features after Feature Engineering:

134



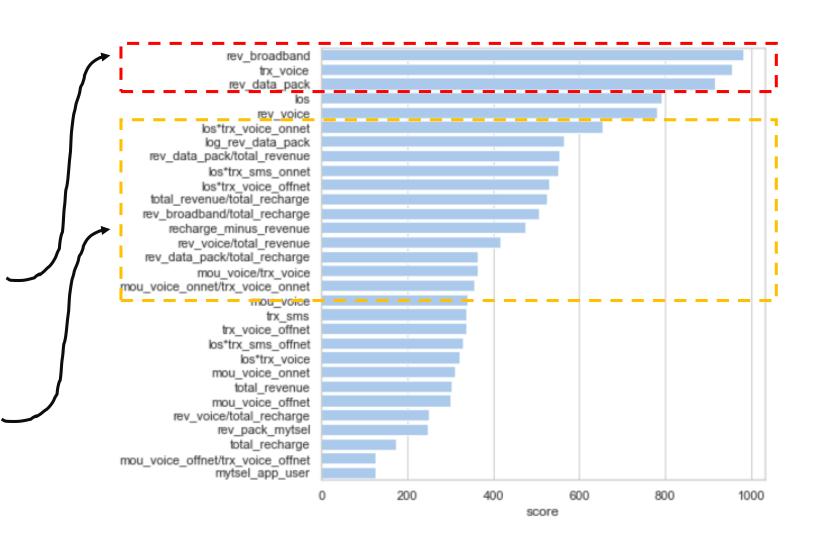




30 top feature ingested into the model, feature selection conducted using XGBoost algorithm

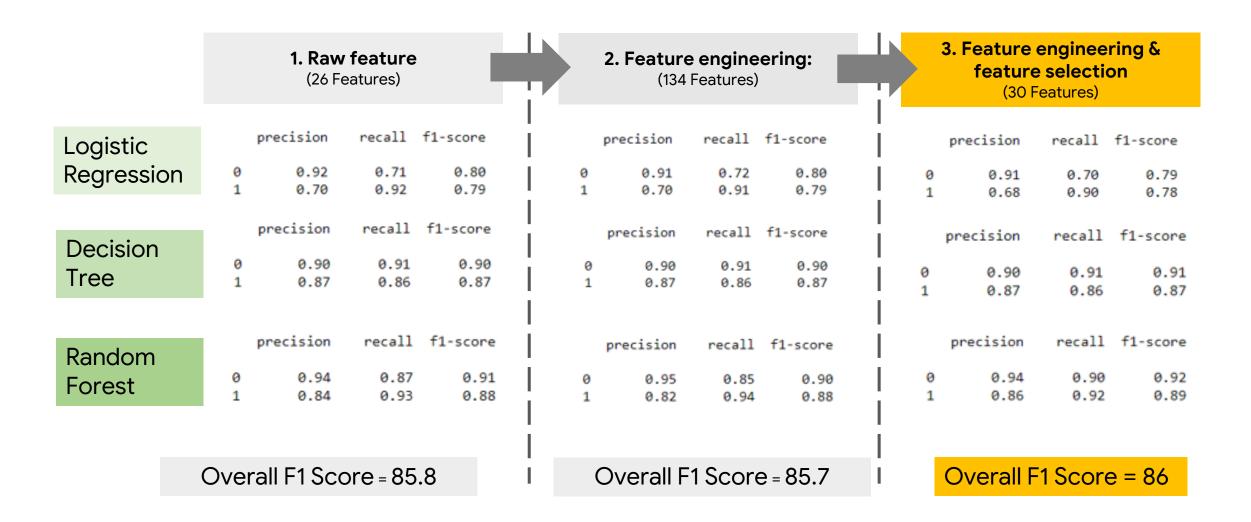
Revenue of data usage (rev_broadband & rev_data_pack) managed to get into top 3

Feature that created from **feature engineering** managed to have high score





To ensure we get the best performing model, we train the model in several iteration. **Iteration 3 have best overall performance (f1 score)** compared to previous iterations.



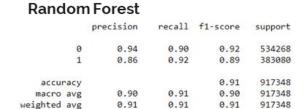


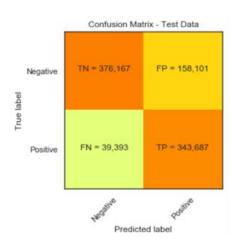
For 1st Objective (Classification), Random Forest achieved highest score, compared to Logistic Regression and Decision Tree

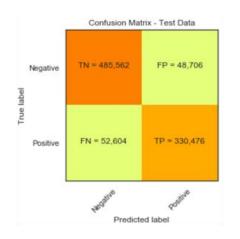
Logistic Regression

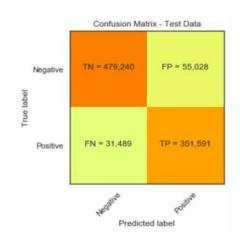
	precision	recall	f1-score	support
0	0.91	0.70	0.79	534268
1	0.68	0.90	0.78	383080
accuracy			0.78	917348
macro avg	0.80	0.80	0.78	917348
weighted avg	0.81	0.78	0.79	917348











F1 Score= 79%

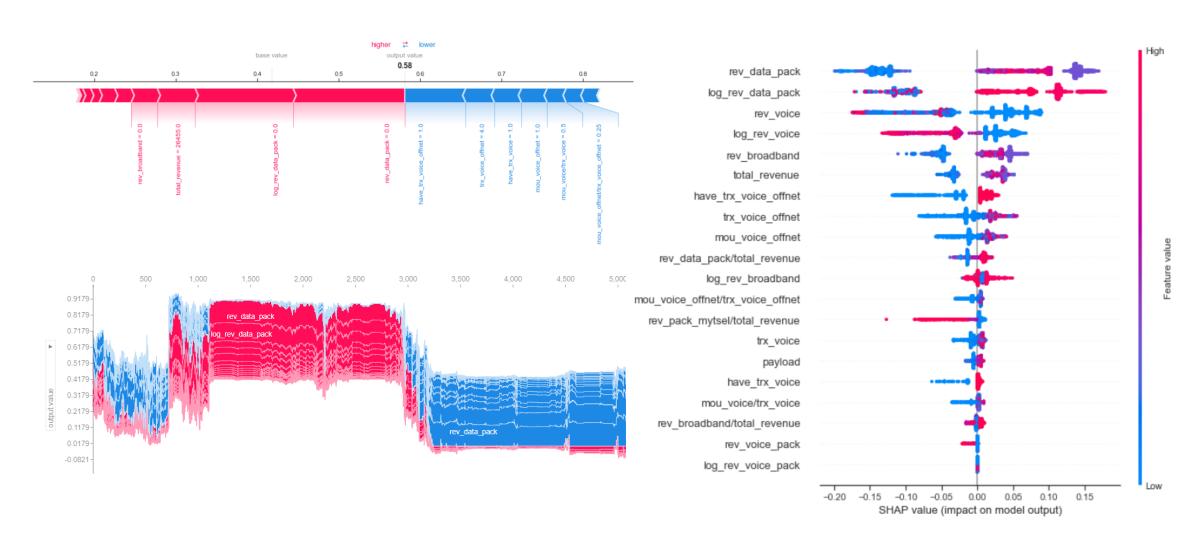
F1 Score= 89%

F1 Score= 91%

All algorithm managed to achieve **key result of F1-Score above 70%**, we decided to went with **Random Forest** that managed to have high precision and recall resulting with **high F1-Score (91%)**

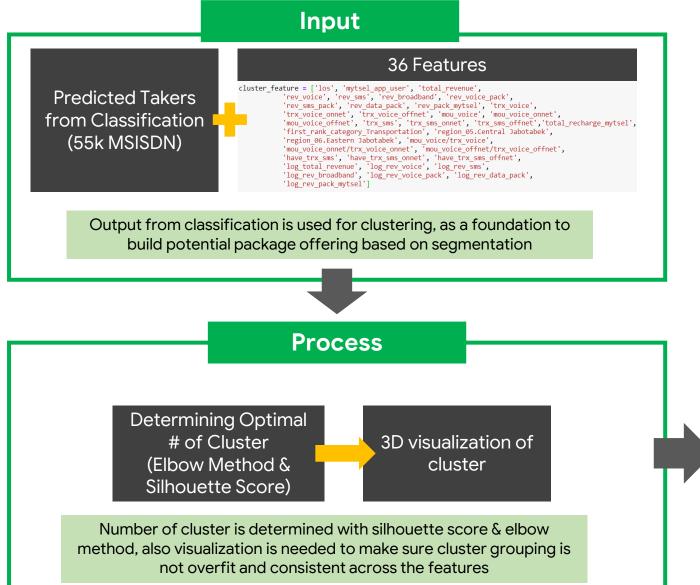


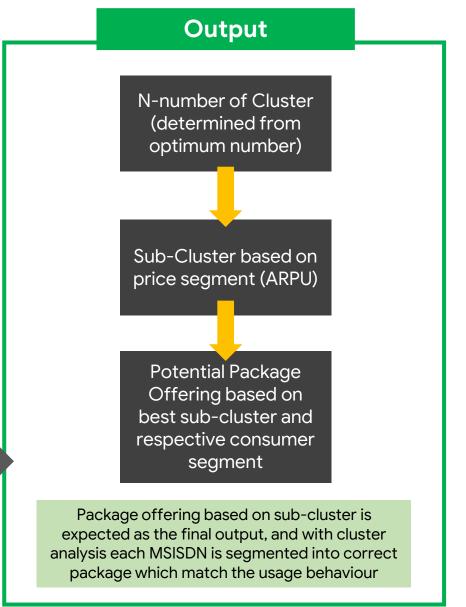
SHAP Value also determined that **data_package** is the most important feature, followed by **voice**





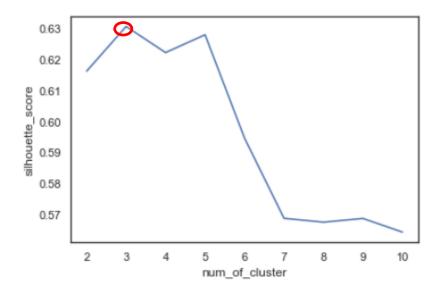
Output from 1st Objective (Classification) will be used for determining cluster and final deliverable for package offering

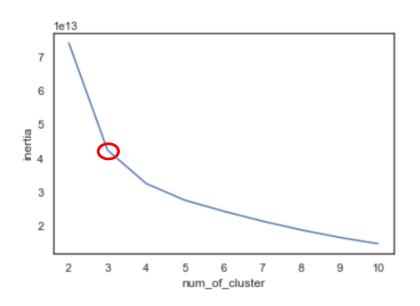






For 2nd Objective (Clustering), silhouette score is used to determine number of cluster. The optimal number of cluster is 3





Determining number of cluster

	num_of_cluster	silhouette_score
0	2	0.616277
1	3	0.630718
2	4	0.622290
3	5	0.628041
4	6	0.594673
5	7	0.568776
6	8	0.567535
7	9	0.568733
8	10	0.564309

Total MSISDN in each cluster

```
pd.DataFrame(cluster, columns=['cluster'])\
['cluster'].value_counts()

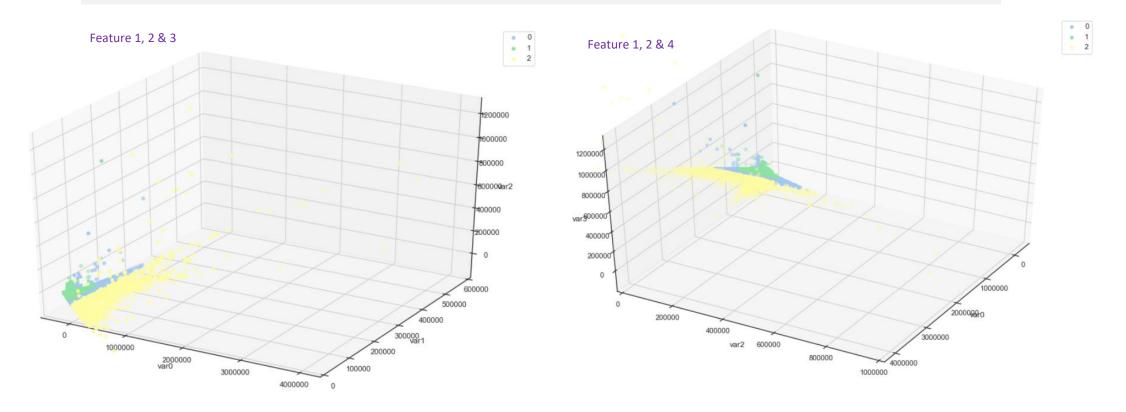
0     46537
2     26699
1     5681
Name: cluster, dtype: int64
```



For 2nd Objective (Clustering), using PCA Analysis and 3D Viz, cluster looks have consistent grouping

3D Cluster Visualization using PCA Analysis Cluster seems have consistent grouping (not underfit/overfit) based on 3D visualization.

Cluster 0 have the largest number of MSISDN (46k), followed by cluster 2 (26k) and cluster 1 (3k)



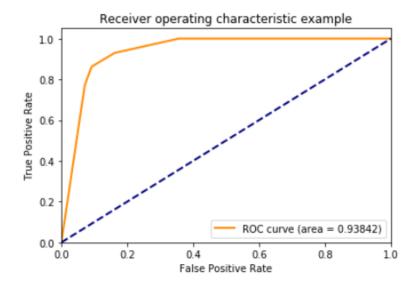


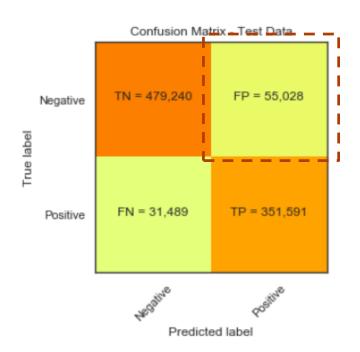




Random Forest successfully meet the 1st objective for classification with 93.8% of AUC and 91% of F1-Score

	precision	recall	f1-score	support
Ø 1	0.94 0.86	0.90 0.92	0.92 0.89	534268 383080
-	0.00	0.52	0.03	303000
accuracy			0.91	917348
macro avg	0.90	0.91	0.90	917348
weighted avg	0.91	0.91	0.91	917348





There are potential **55k new numbers of takers** based on this algorithm. In total there are **406k potential package takers.**

44.2% taker rate, uplift +4% from previous data



Data

Addict

253k

51 GB

139 Min

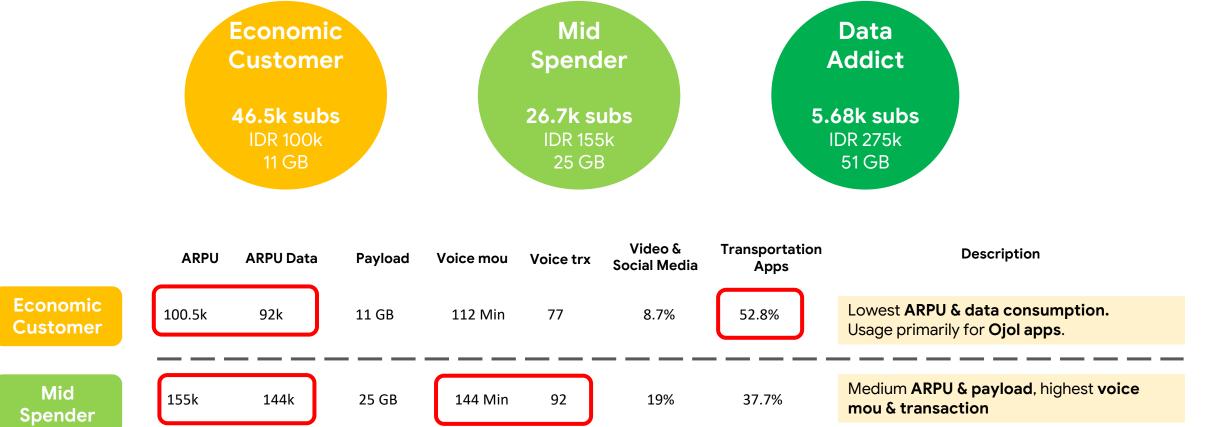
84

43.7%

18.2%

275k

For the 2nd Objective, there are 3 main cluster with different behaviour and usage, resulting in different package price



High data consumption & dominant video

& social app usage.



For the 3rd Objective, main cluster derived into each 3 sub-cluster each, resulting in 9 different sub-cluster

	Lower ARPU (< IDR 75k)	Middle ARPU (IDR 75k - IDR 150k)	Top ARPU (>IDR 150k)
Economic Customer	3.9k subs IDR 39k 9 GB	35.8k subs IDR 90k 11 GB	6.8k subs IDR 188k 13 GB
Mid Spender	405 subs IDR 45k 23 GB	10.8k subs IDR 99k 23 GB	15.5k subs IDR 198k 27 GB
Data Addict	38 subs IDR 42k 50 GB	306 subs IDR 117k 49 GB	5.3k subs IDR 285k 51 GB

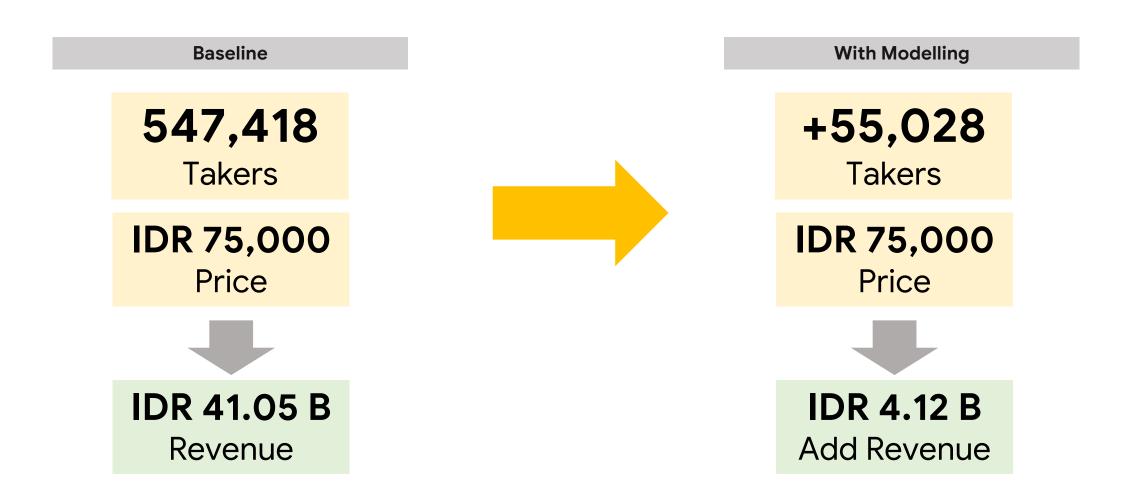
Cluster	Subsegment	count msisdn	total revenue	rev broadband	payload	trx voice	mou voice	video & social media	first rank category Transportation
	< 75k	3,905	39,434	37,079	9,390,430	67	91	7.810%	54.366%
Economic Customer	Between 75k and 150k	35,807	90,432	84,548	11,666,051	79	116	9.051%	52.400%
	> 150k	6,825	188,698	166,820	13,440,510	71	101	7.897%	54.505%
	< 75k	405	44,987	41,941	23,055,692	79	137	17.284%	32.840%
Mid Spender	Between 75k and 150k	10,794	98,828	92,451	23,460,466	92	146	17.612%	38.614%
	> 150k	15,500	198,487	183,510	27,443,753	92	142	20.213%	37.200%
	< 75k	38	42,810	39,188	50,200,170	111	159	57.895%	10.526%
Data Addict	Between 75k and 150k	306	117,133	105,623	49,436,040	82	108	41.503%	18.301%
	> 150k	5,337	285,669	263,442	51,643,722	84	141	43.676%	18.269%





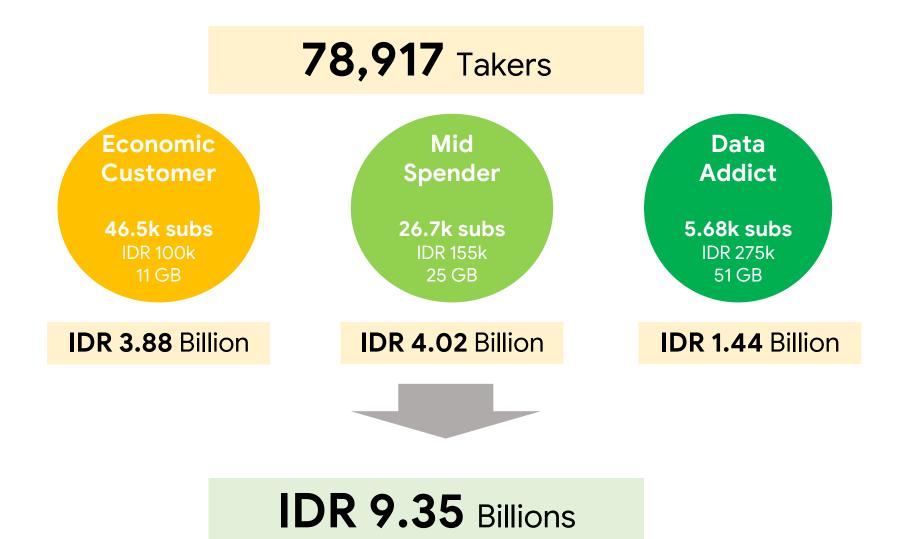


Revenue projection from 1st Objective can be seen from predicted takers, that add **4.12 billion rupiah** of additional revenue



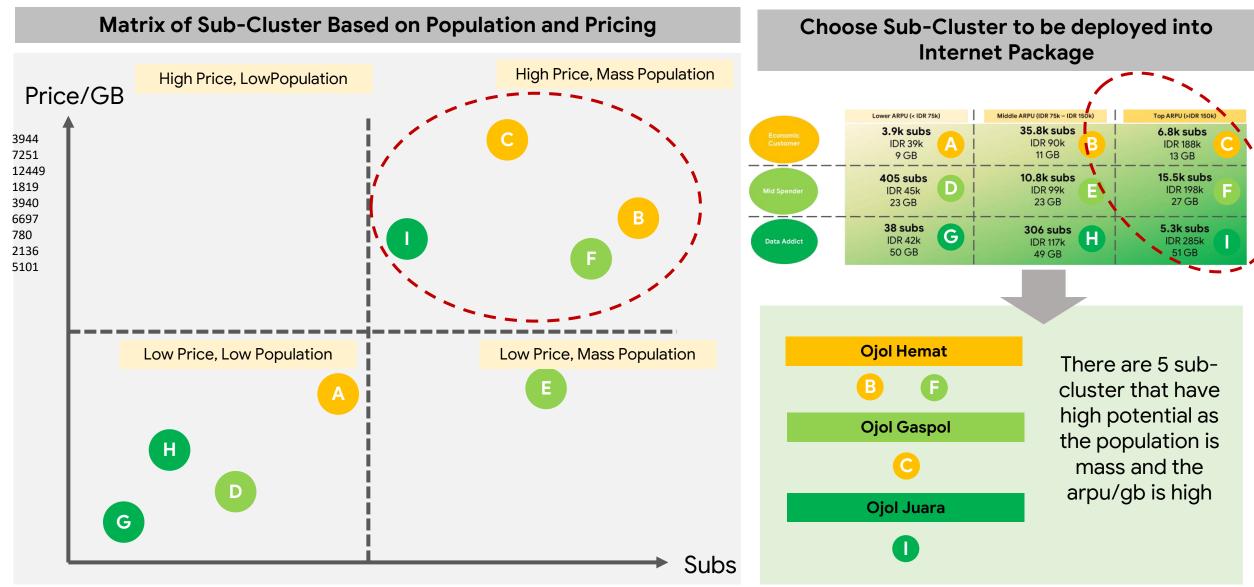


For the 2nd Objective with changes in price offering based on ARPU we can even get additional revenue up to **IDR 9.35 Billions**





For the deployment of 3rd objective, there are **5 sub-cluster selected based on price and population size**





Our package offering based on sub-cluster splitted into 4 different package that can bring home uplift of IDR 5320/subs

Current Telkomsel Offering

Baseline

Tipe Layanan	Paket Swadaya Telkomsel
Biaya per bulan	Rp75.000
Masa periode aktif	30 Hari
Kuota telepon ke sesama operator	Tidak terbatas atau unlimited
Kuota telepon ke semua operator	200 menit
SMS	500 SMS
Kuota internet	15 GB

Takers **78,197**

Price **IDR 75,000**

Revenue IDR 5,864,775,000

Ojol Hemat

	Layanan	Harga
^	Harga	Rp80.000
A	Kuota	20 GB
	SMS	50 SMS
	Telepon	50 Menit
	Telepon	50 Menit

Taker 39712 Revenue IDR 3,17 B

	Layanan	Harga
D	Harga	Rp150.000
В	Kuota	25 GB
	SMS	100 SMS
	Telepon	100 Menit

Taker **26699**

Revenue IDR 4,00 B

Ojol Gaspol

Layanan	Harga
Harga	Rp175.000
Kuota	30 GB
SMS	200 SMS
Telepon	200 Menit

Taker **6825**

Revenue IDR 1,19 B

Ojol Gaspol

Layanan	Harga
Harga	Rp255.000
Kuota	50 GB
SMS	200 SMS
Telepon	200 Menit

Taker **5681**

Revenue IDR 1,44 B

		Data Revenue	Data Usage (Byte)	SMS Trx	Voice Minutes	Subs	Total Revenue	
ol Hemat A	76,853.78	71,115.39	12,183,935.62	42.10	45.51	39,712.00	3052017474	
ol Gaspol	168,296	152,394	14,670,167	34	34	6,825	1148620200	
ol Hemat B	141,800.13	130,011.77	24,931,172.58	53.45	65.21	26,699	3785921704	
ol Juara	252,491.96	235,912.93	47,044,447.69	47.08	60.04	5,681	1434406821	
				SUM		78,917.00	9420966199	
ol ol	Gaspol Hemat B	Gaspol 168,296 Hemat B 141,800.13	Gaspol 168,296 152,394 Hemat B 141,800.13 130,011.77	Gaspol 168,296 152,394 14,670,167 Hemat B 141,800.13 130,011.77 24,931,172.58	Gaspol 168,296 152,394 14,670,167 34 Hemat B 141,800.13 130,011.77 24,931,172.58 53.45 Juara 252,491.96 235,912.93 47,044,447.69 47.08	Gaspol 168,296 152,394 14,670,167 34 34 Hemat B 141,800.13 130,011.77 24,931,172.58 53.45 65.21 Juara 252,491.96 235,912.93 47,044,447.69 47.08 60.04	Gaspol 168,296 152,394 14,670,167 34 34 6,825 Hemat B 141,800.13 130,011.77 24,931,172.58 53.45 65.21 26,699 Juara 252,491.96 235,912.93 47,044,447.69 47.08 60.04 5,681	

	Paket	Price	Data Usage (GB)	SMS Trx	Voice Minutes	Subs	Total Revenue	Add Revenue	Uplift/Subs
Package	Ojol Hemat A	80000	20 GB	50	50	39712	3176960000	124942526	3146.215905
Pricing	Ojol Gaspol	175000	30 GB	200	200	6825	1194375000	45754800	6704
Fricing	Ojol Hemat B	150000	25 GB	100	100	26699	4004850000	218928296	8199.868759
	Ojol Juara	255000	50 GB	200	200	5681	1448655000	14248179	2508.040662
					SUM	78917	9824840000	389625622	4937.157038

Total Revenue **IDR 9.824.840.000**

+67%

Uplift/Subs IDR 4937

A	ט	C	U	L	ı	0
	•	. T	los 🔻	total_recharge 🔻	mytsel_app_user 🔻	total_revenue 🔻 rev_
		count	46,537	46,537	46,537	46,537
ow Payload,	Economic Customer	mean	1,428	98,538	44%	100,564
		50%	896	85,000	0%	83,551
		count	26,699	26,699	26,699	26,699
ligh voice mo	Traditional Customer	mean	1,366	152,634	55%	155,868
		50%	869	150,000	100%	150,680
		count	5,681	5,681	5,681	5,681
ligh Payload,	Data Addict	mean	1,367	265,238	67%	274,966
		50%	948	250,000	100%	254,981
					78,917	3888212887
						4,023,005,320
						1448547061
						9359765268

	Paket	Revenue	Data Revenue	Data Usage (Byte)	SMS Trx	Voice Minutes	Subs	Total Revenue
Sub A & B	Ojol Hemat A	76,853.78	71,115.39	12,183,935.62	42.10	45.51	39,712.00	3052017474
Sub C	Ojol Gaspol	168,296	152,394	14,670,167	34	34	6,825	1148620200
Sub D, E, F	Ojol Hemat B	141,800.13	130,011.77	24,931,172.58	53.45	65.21	26,699	3785921704
Sub G, H, I	Ojol Juara	252,491.96	235,912.93	47,044,447.69	47.08	60.04	5,681	1434406821
						SUM	78,917.00	9420966199

	Paket	Price	Data Usage (GB)	SMS Trx	Voice Minutes	Subs	Total Revenue	Add Revenue	Uplift/Subs
Daaliaaa	Ojol Hemat A	80000	20 GB	50	50	39712	3176960000	124942526	3146.215905
Package Pricing	Ojol Gaspol	175000	30 GB	200	200	6825	1194375000	45754800	6704
Fricing	Ojol Hemat B	150000	25 GB	100	100	26699	4004850000	218928296	8199.868759
	Ojol Juara	255000	50 GB	200	200	5681	1448655000	14248179	2508.040662
					SUM	78917	9824840000	389625622	4937.157038

GO-JEK berkontribusi Rp 8,2 triliun per tahun ke dalam perekonomian Indonesia melalui penghasilan Mitra Pengemudi.



		Sebelum	Setelah menjadi mitra			Total Pendapatan yang masuk dalam perekonomian					
Penghasilan Sebelum menjadi Mitra				Jumlah	Jumlah		Jumlah	bulan (Ribu Rupiah)			
		Jumlah Responden (Survei)	Proporsi	Responden Weighted**	Responden (Survei)	Proporsi	Responden Weighted**	Total Sebelum	Total Sesudah	Selisih	
<1 juta	500	133	4%	27,081	39	1%	7,941	13,540,723.98	3,970,588.24		
1-1,5 juta	1250	302	9%	61,493	203	6%	41,335	76,866,515.84	51,668,552.04		
1,5-2 juta	1,750	707	21%	143,959	296	9%	60,271	251,928,733.03	105,475,113.12		
2-2,5 juta	2,250	982	30%	199,955	475	14%	96,719	449,898,190.05	217,618,778.28		
2,5-3,5 juta	3,000	799	24%	162,692	1148	35%	233,756	488,076,923.08	701,266,968.33		
3,5-6 juta	4,750	213	6%	43,371	1041	31%	211,968	206,012,443.44	1,006,849,547.51		
>6 juta	6,500	51	2%	10,385	113	3%	23,009	67,500,000.00	149,558,823.53		
Tidak Bekerja Sebelumnya	-	128	4%	26,063	0	0	-	-	-		
Total		3315	100%	675,000*	3315	100%	675,000	1,553,823,529.41	2,236,408,371.04	682,584,841	

^{*}http://tekno.kompas.com/read/2017/12/18/07092867/berapa-jumlah-pengguna-dan-pengemudi-GO-JEK

^{**}Weight berasal dari hasil survei yang telah diolah