

Auto Insurance analysis and Churn prediction Project

Presented by: Aditya Aji Nugroho

Overview

- Exploratory Data Analysis (EDA)
- Analysis
- Machine Learning
- Business insight



Background

Dataset ini memiliki sekitar 1.62 juta rows dan 22 kolom didalamnya. Dataset ini terfokus pada infromasi para customer yang sudah menjadi anggota pada Auto_insurance ini. Tujuan dari presentasi ini adalah untuk memperlihatkan bagaimana kita dapat memberikan business insight dan saran menggunakan data yang ada .

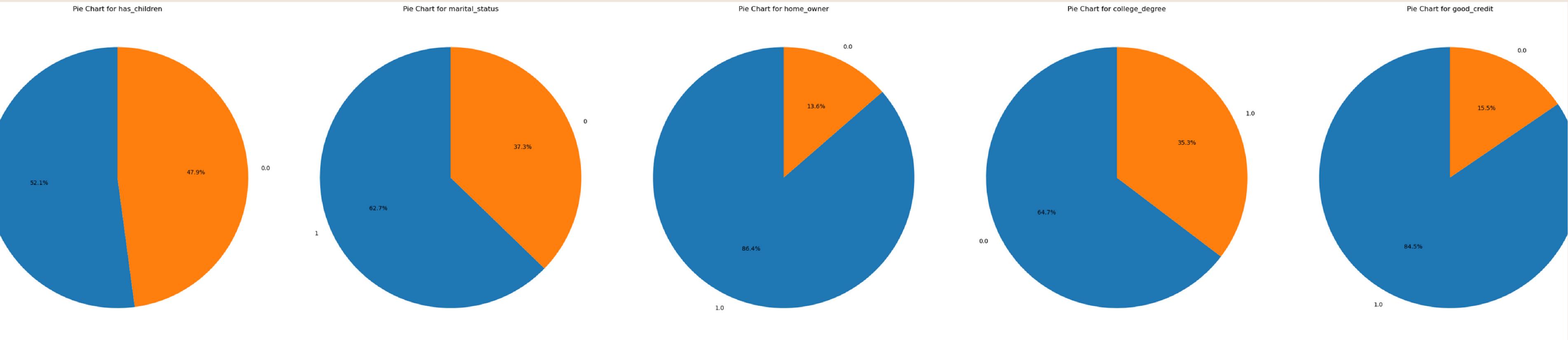
Handling Missing values

	feature	missing_value	percentage
0	acct_suspd_date	1487453	88.49
1	latitude	253719	15.09
2	longitude	253719	15.09
3	home_market_value	92286	5.49
4	county	12067	0.72
5	city	12067	0.72

Handling Missing Values

- Saya memutuskan untuk melakukan drop terhadap Acct_suspd karena memiliki redundancy terhadap kolom Churn
- Untuk Longtitude dan Latitude saya memutuskan melakukan drop untuk kedua kolom karena hanya menyertakan kordinat dan tidak memberikan pengaruh untuk analisis
- Saya mendrop county dan city karena tidak memiliki sentiment analisis dan tidak memiliki pengaruh untuk analisis
- Saya memutuskan untuk mempertahankan home_market_value karena memiliki pengaruh untuk analisis

Proportion of columns

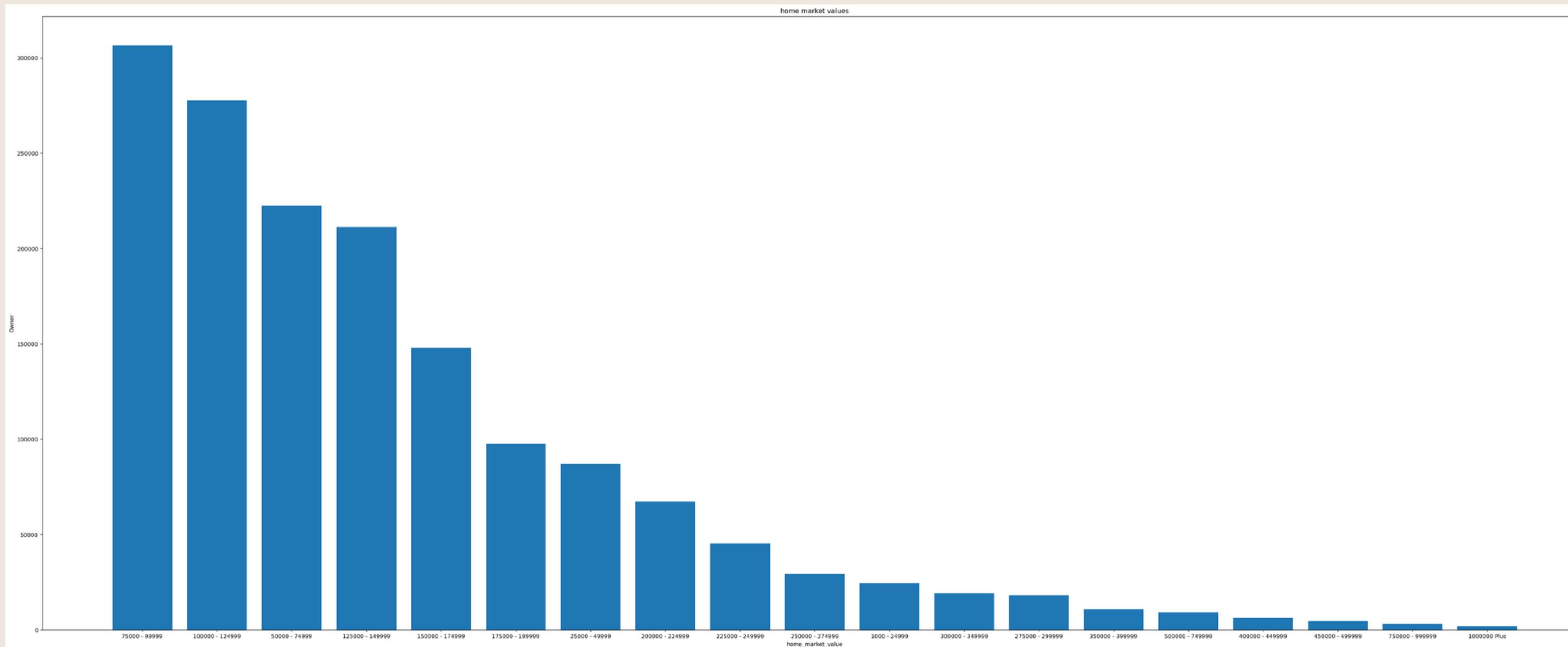


- Mayoritas customer memiliki anak 52.1% punya anak dan 47.9% tidak.
- Mayoritas customer sudah menikah 62.7% menikah dan 37.3% belum.
- Mayoritas customer adalah pemilik rumah 86.4% memiliki rumah dan 13.6% mengontrak
- Mayoritas customer memiliki ijasah kuliah 64.7% kuliah dan 35.3% tidak kuliah
- Mayoritas customer memiliki credit score yang baik 84.5% baik dan 15.5% buruk.
- Mayoritas customer masih menggunakan produk perusahaan 88.6% dan 12.4% sudah Churn.

Home Ownership Analysis

	home_market_value	total_customer	home_ownership	resident_lenght	status	cust_income
0	1000 - 24999	17847	12903.0	7.141230	Single	70000.0
1	100000 - 124999	255216	232346.0	8.128942	Married	87500.0
2	1000000 Plus	1399	1166.0	6.166569	Married	250000.0
3	125000 - 149999	190259	173966.0	7.843116	Married	87500.0
4	150000 - 174999	128934	117774.0	7.520063	Married	125000.0
5	175000 - 199999	83602	77022.0	7.185494	Married	125000.0
6	200000 - 224999	56411	51792.0	6.959408	Married	125000.0
7	225000 - 249999	37011	33845.0	6.759853	Married	125000.0
8	25000 - 49999	79081	69042.0	9.271681	Married	32500.0
9	250000 - 274999	22663	20095.0	6.583746	Married	125000.0
10	275000 - 299999	13894	12647.0	6.837309	Married	250000.0
11	300000 - 349999	13008	11215.0	6.929649	Married	250000.0
12	350000 - 399999	7053	5996.0	6.552297	Married	125000.0
13	400000 - 449999	4053	3476.0	6.321660	Married	125000.0
14	450000 - 499999	2903	2440.0	6.446524	Married	125000.0
15	50000 - 74999	205842	185818.0	9.274584	Married	87500.0
16	500000 - 749999	5894	4831.0	6.638300	Married	125000.0
17	75000 - 99999	282199	256286.0	8.598831	Married	87500.0
18	750000 - 999999	2242	1941.0	6.877364	Married	250000.0

Customer Houses Values

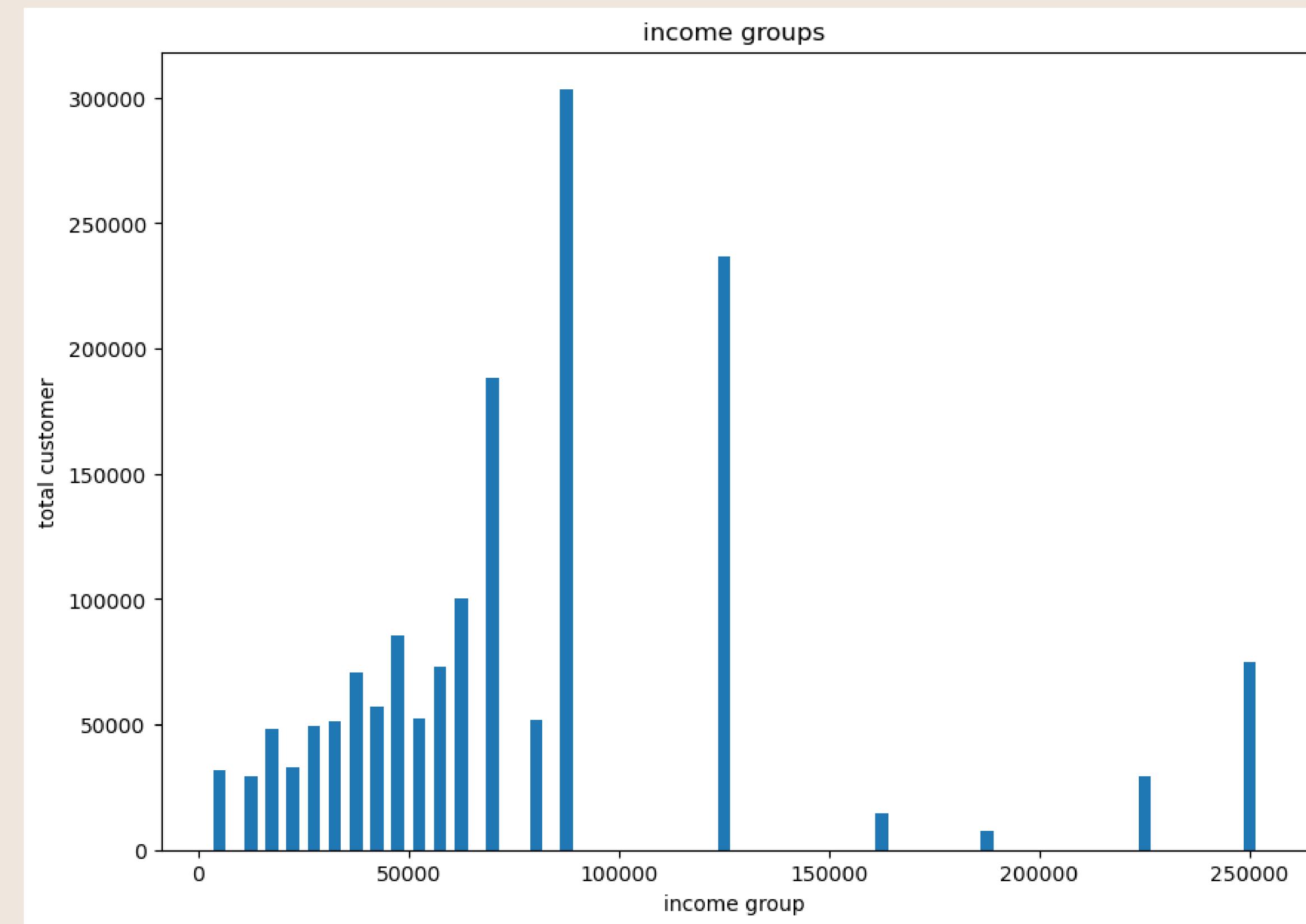


Customer lebih banyak tinggal di Rumah dengan
harga range 75000 - 99999

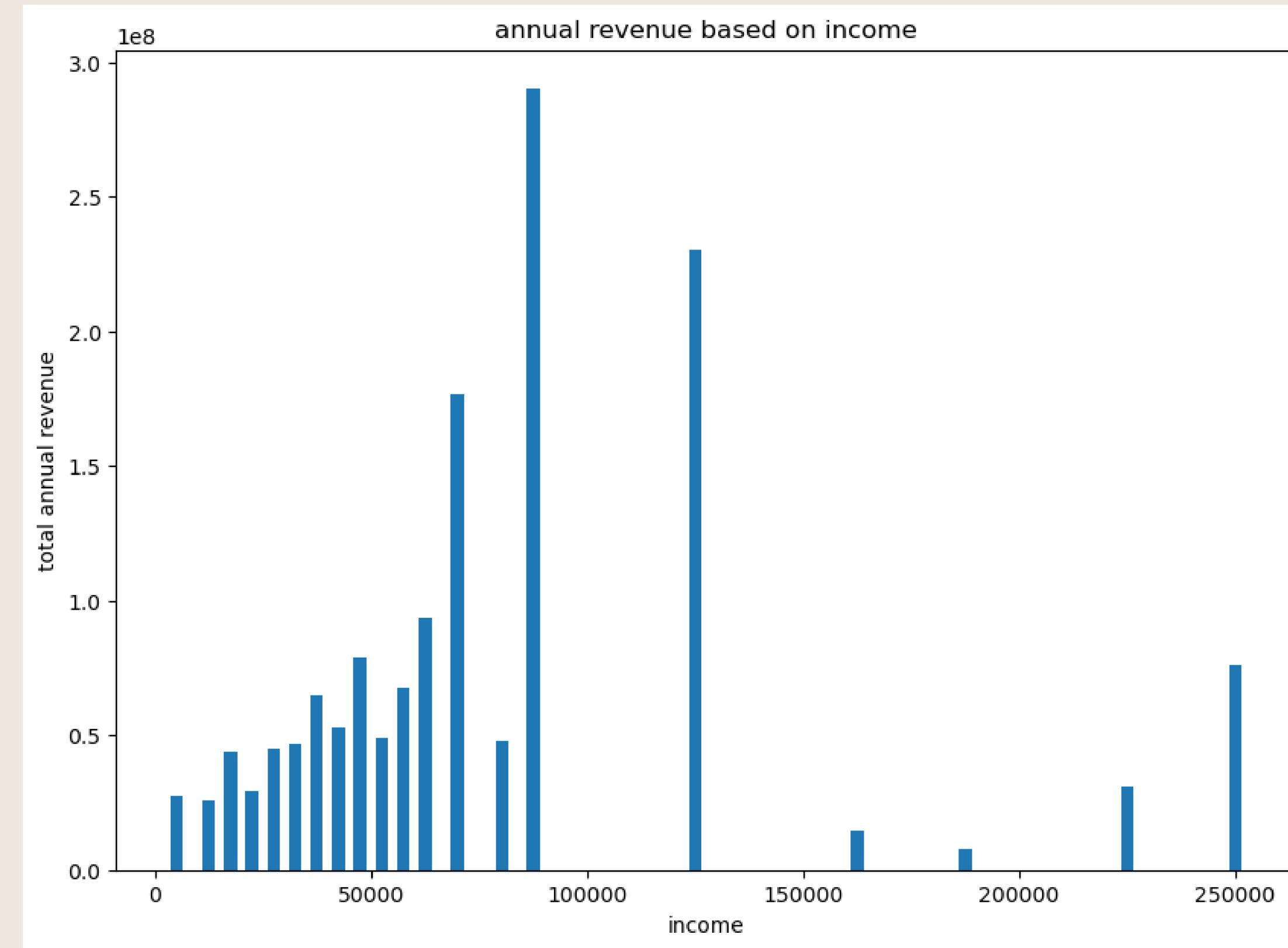
Customer Analysis

	income	total_customer	mean_age	annual_revenue	good_credits	have_degree	Churn_total
0	5000.000	31721	59.817282	2.772363e+07	3732.0	10333.0	3886
1	12500.000	29328	59.411654	2.594315e+07	8720.0	10427.0	3424
2	17500.000	48525	61.894611	4.372035e+07	41405.0	15803.0	5493
3	22500.000	32622	56.855404	2.927681e+07	8088.0	10068.0	3823
4	27500.000	49672	59.417257	4.488042e+07	43734.0	14850.0	5501
5	32500.000	51132	55.705938	4.665713e+07	20295.0	15495.0	5977
6	37500.000	70883	57.640266	6.494828e+07	64697.0	20845.0	8216
7	42500.000	56920	55.415039	5.269212e+07	35893.0	17293.0	6965
8	47500.000	85569	55.760941	7.872951e+07	79366.0	25228.0	9781
9	52500.000	52702	55.535160	4.925802e+07	38514.0	16233.0	5944
10	57500.000	72836	55.032333	6.776090e+07	67943.0	21730.0	8258
11	62500.000	100447	54.885552	9.387470e+07	90457.0	31283.0	11423
12	70000.000	188438	55.174970	1.766402e+08	173762.0	63555.0	21553
13	80372.176	51660	50.572493	4.777300e+07	40365.0	10339.0	6146
14	87500.000	303226	55.245464	2.900244e+08	279834.0	111296.0	34617
15	125000.000	236751	56.093136	2.304409e+08	223884.0	102221.0	26489
16	162500.000	14562	56.965046	1.436751e+07	13471.0	6721.0	1639
17	187500.000	7618	57.360331	7.681658e+06	7201.0	3587.0	808
18	225000.000	29379	57.574764	3.093385e+07	28545.0	14615.0	3182
19	250000.000	74632	57.463581	7.628525e+07	73106.0	39462.0	8235

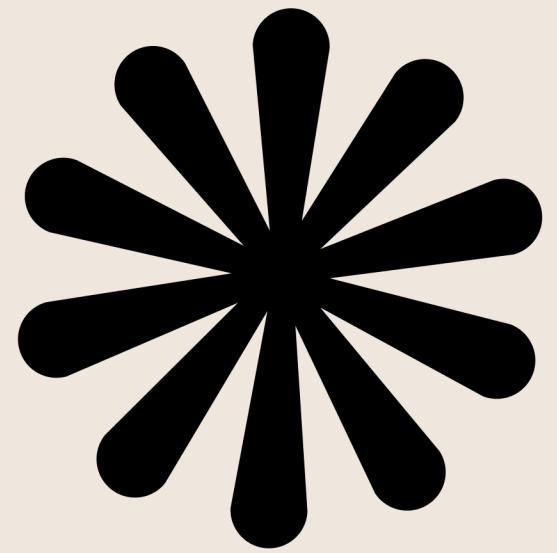
Customer Income Groups



Annual Revenue per income groups



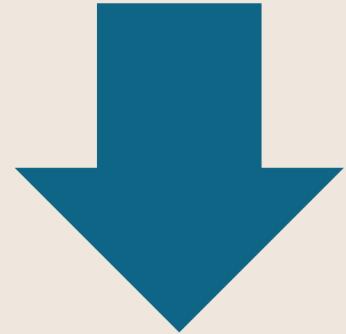
kelompok income 87500 memiliki total revenue value tersebesar yaitu 290 juta



Machine Learning

Under Sampling

```
Value counts of Churn column  
0    1487453  
1    193456  
Name: Churn, dtype: int64
```



```
Churn  
0    159818  
1    159818  
dtype: int64
```

RFE

1	Logistic Regression:												
2		mean_te st_score	std_test_ score	split0_te st_score	split1_te st_score	split2_te st_score	split3_te st_score	split4_te st_score	split5_te st_score	split6_te st_score	split7_te st_score	split8_te st_score	split9_te st_score
3	0	0.652276	0.003136	0.654768	0.652272	0.652293	0.649797	0.653083	0.653894	0.644596	0.657139	0.652396	0.652521
4	1	0.648931	0.003275	0.651086	0.650087	0.648362	0.646802	0.649713	0.649089	0.640851	0.654393	0.649214	0.649713
5	2	0.648438	0.002925	0.649775	0.650150	0.647613	0.645803	0.649401	0.649276	0.641226	0.652646	0.648714	0.649775
6	3	0.648376	0.002794	0.650524	0.649651	0.647551	0.646303	0.648777	0.649276	0.641350	0.652334	0.648777	0.649214
7	4	0.648114	0.003039	0.649963	0.649651	0.646989	0.645741	0.648090	0.649089	0.640789	0.652958	0.649276	0.648590
8	5	0.647845	0.002985	0.649151	0.648902	0.646927	0.646115	0.648215	0.649151	0.640352	0.652708	0.648402	0.648527
9	6	0.647458	0.003193	0.649214	0.648777	0.645803	0.645367	0.648028	0.648402	0.639978	0.653145	0.647778	0.648090
10	7	0.647533	0.003197	0.649338	0.649463	0.646178	0.645616	0.647654	0.648215	0.639978	0.653145	0.647404	0.648340
11	8	0.647571	0.003312	0.649276	0.649588	0.646178	0.645616	0.647591	0.648527	0.639666	0.653333	0.647778	0.648153
12	9	0.647583	0.003340	0.649276	0.649526	0.645928	0.645679	0.648028	0.648340	0.639541	0.653270	0.647841	0.648402
13	10	0.647564	0.003249	0.648902	0.649338	0.645991	0.645616	0.648028	0.648402	0.639853	0.653333	0.647841	0.648340
14	11	0.647508	0.003202	0.648839	0.649338	0.645991	0.645679	0.647778	0.648215	0.639853	0.653145	0.648028	0.648215
15	12	0.647489	0.003197	0.648714	0.649214	0.646053	0.645554	0.647841	0.648402	0.639853	0.653145	0.647966	0.648153
16	13	0.647446	0.003163	0.648527	0.649276	0.646178	0.645491	0.647966	0.648590	0.639853	0.652958	0.647466	0.648153
17	14	0.647471	0.003085	0.648902	0.649026	0.645991	0.645554	0.647966	0.648777	0.640040	0.652646	0.647404	0.648402
18	15	0.647645	0.003037	0.648714	0.649276	0.646490	0.646365	0.647903	0.648777	0.640102	0.652771	0.647466	0.648590
19	16	0.643371	0.006113	0.648527	0.639915	0.635257	0.635944	0.637731	0.648028	0.638729	0.652771	0.648153	0.648652
20	17	0.636662	0.003488	0.637793	0.639915	0.635320	0.635757	0.637731	0.635609	0.628183	0.641226	0.635235	0.639853
21	18	0.635994	0.003567	0.639666	0.638230	0.633573	0.635881	0.637294	0.635796	0.627371	0.641101	0.635172	0.635859
22	19	0.635982	0.003593	0.639666	0.638168	0.633822	0.635881	0.637294	0.635671	0.627184	0.641101	0.635172	0.635859

1	xgBoost:												
2		mean_te st_score	std_test_ score	split0_te st_score	split1_te st_score	split2_te st_score	split3_te st_score	split4_te st_score	split5_te st_score	split6_te st_score	split7_te st_score	split8_te st_score	split9_te st_score
3	0	0.473646	0.003822	0.475037	0.478844	0.468456	0.472574	0.473290	0.468922	0.469608	0.477908	0.472604	0.479219
4	1	0.473116	0.004084	0.474164	0.478782	0.467083	0.471451	0.472916	0.469109	0.468672	0.476910	0.472354	0.479718
5	2	0.456297	0.003991	0.458063	0.460809	0.450546	0.455101	0.456628	0.451323	0.451885	0.459873	0.455816	0.462931
6	3	0.451174	0.003607	0.451448	0.455192	0.445616	0.449548	0.451760	0.447391	0.447641	0.454568	0.450949	0.457626
7	4	0.449620	0.003795	0.450574	0.453819	0.444555	0.447176	0.448952	0.444458	0.446705	0.453445	0.450325	0.456191
8	5	0.449751	0.003834	0.450387	0.452883	0.444618	0.447925	0.448640	0.445457	0.447017	0.453694	0.449139	0.457751
9	6	0.453914	0.004245	0.456003	0.457564	0.448674	0.450858	0.455005	0.448078	0.449825	0.457938	0.453694	0.461495
10	7	0.454176	0.004063	0.455816	0.456690	0.449485	0.451357	0.454693	0.448328	0.449950	0.458937	0.455255	0.461246
11	8	0.455461	0.004381	0.457189	0.458437	0.451045	0.451232	0.456128	0.448203	0.452384	0.460185	0.456877	0.462931
12	9	0.456560	0.004539	0.456628	0.460185	0.450858	0.453292	0.456815	0.450325	0.452571	0.463555	0.458250	0.463118
13	10	0.456709	0.004216	0.457252	0.461745	0.450608	0.452668	0.456815	0.451573	0.453819	0.462307	0.458063	0.462244
14	11	0.457265	0.003766	0.457439	0.462369	0.452418	0.455101	0.456690	0.451947	0.454818	0.462119	0.457064	0.462681
15	12	0.457408	0.003844	0.458562	0.460497	0.452730	0.455226	0.457189	0.451760	0.454443	0.462182	0.457252	0.464241
16	13	0.457776	0.004200	0.461183	0.460497	0.453354	0.454977	0.458250	0.451635	0.453008	0.462431	0.457751	0.464678
17	14	0.458600	0.004327	0.461246	0.462619	0.452605	0.454727	0.458000	0.453944	0.455067	0.463929	0.458188	0.465676
18	15	0.458800	0.003930	0.461246	0.463367	0.452480	0.457036	0.458874	0.453320	0.456128	0.461807	0.458749	0.464990
19	16	0.458388	0.003914	0.461807	0.462369	0.452543	0.456412	0.457439					

Modelling

	method	accuracy	auc_score	precision	recall	f1_score
0	DecisionTreeClassifier	0.592136	0.592141	0.590813	0.595729	0.593261
1	XGBClassifier	0.690485	0.690171	0.846762	0.464097	0.599576
2	RandomForestClassifier	0.681851	0.681585	0.793798	0.490138	0.606059
3	LogisticRegression	0.625249	0.625284	0.618542	0.650818	0.634270

RandomizeCV

	precision	recall	f1-score	support
0	0.63	0.60	0.62	40010
1	0.62	0.65	0.64	39899
accuracy			0.63	79909
macro avg	0.63	0.63	0.63	79909
weighted avg	0.63	0.63	0.63	79909

Business Insight

- Revenue terbanyak berasal dari kelompok income 87500
- Total Churn terbanyak juga berasal dari kelompok income 87500
- Kebanyakan customer tinggal di rumah yang memiliki range value sebesar 75000 dengan sekitar 300000 yang tinggal dan 271000 yang memiliki rumah tersebut dengan rata-rata tinggal 8 tahun
- Setelah dilakukan beberapa test dengan beberapa model didapatkan bahwa logistic regression memiliki recall score tertinggi yaitu 65%.

Suggestion:

- Perusahaan dapat fokus untuk menjual produk ke kelompok yang memiliki Income 87500 karena dari data yang sudah diambil kelompok ini menunjukkan kestabilan ekonomi dan memiliki potensi untuk keuntungan yang besar.
- Untuk Churn tidak ada urgensi untuk membuat keputusan, namun dapat memberikan treatment khusus kepada kelompok income 87500 seperti pemberian proteksi khusus atau pemotongan biaya permi.