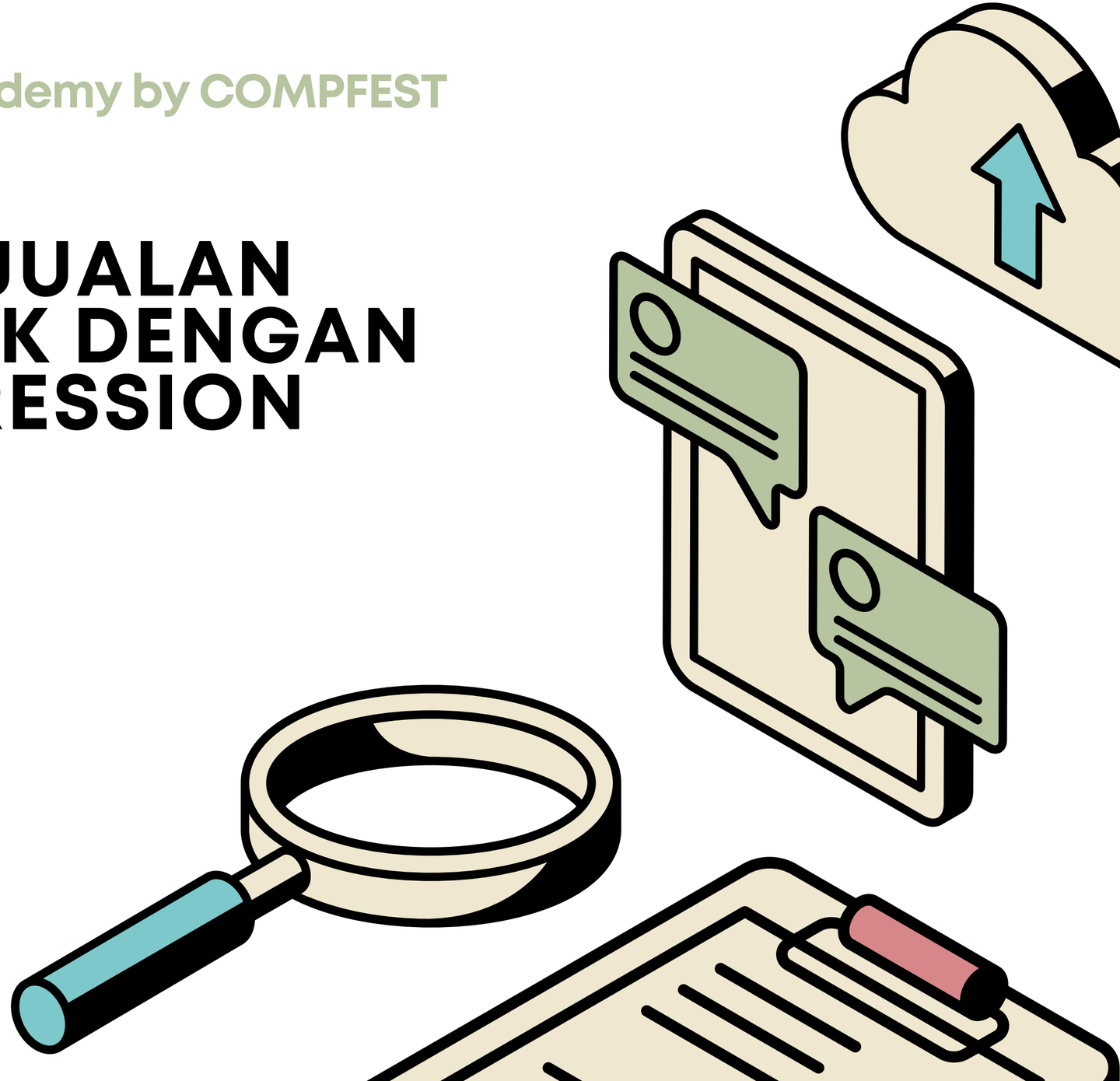
Final Project for Data Science Academy by COMPFEST

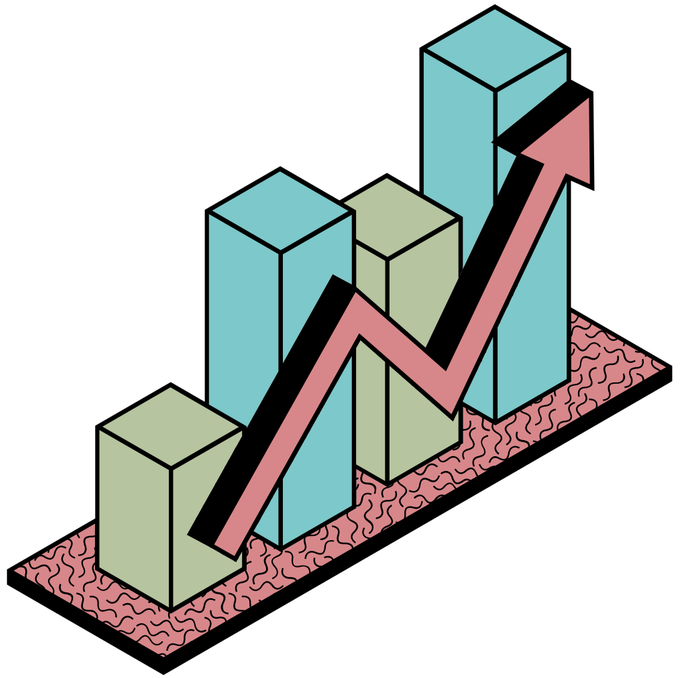
MEMPREDIKSI PENJUALAN KENDARAAN LISTRIK DENGAN POLYNOMIAL REGRESSION



POTTERHEADS

- Deanarani Kharisma
- Romauli Graciella Debora
- Zalfa Aretha Fahira Ranstya

Daftar Isi



Latar Belakang

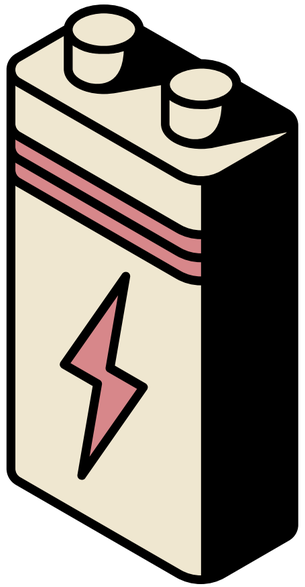
Kendaraan listrik telah mengalami pertumbuhan pesat dalam beberapa tahun terakhir sebagai alternatif ramah lingkungan dalam industri otomotif. Presentasi ini mengulas perkembangan teknologi kendaraan listrik serta analisis produksi total kendaraan listrik per tahun. Dengan fokus pada aspek manufaktur, project ini membahas peningkatan signifikan dalam produksi kendaraan listrik di Negara Washington. Data statistik mengenai total produksi kendaraan listrik per tahun dianalisis untuk mengidentifikasi tren pertumbuhan dan dampaknya terhadap industri otomotif konvensional.

Rumusan Masalah

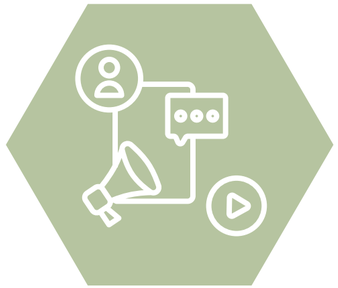
1. Bagaimana tren pertumbuhan produksi kendaraan listrik?

2. Bagaimana akurasi prediksi model machine learning?

3. Bagaimana hasil analisis terhadap model yang diperoleh?



Manfaat Hasil Prediksi



Strategi Pemasaran

Mengarahkan sumber daya pemasaran ke wilayah atau segmen pasar yang memiliki potensi pertumbuhan lebih tinggi.

Pengembangan Produk

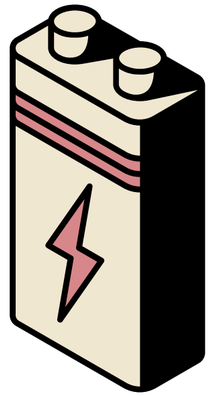
Dapat menghemat biaya produksi dan mengoptimalkan pengelolaan persediaan.

Analisis Persaingan

Dapat membantu dalam merumuskan strategi kompetitif.

Pertimbangan Investasi

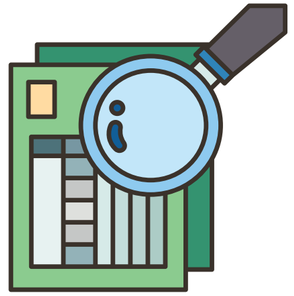
Membantu investor dalam pengambilan keputusan investasi dalam perusahaan otomotif.



Data Pre-processing & Cleaning

Data Observation

HTTPS://CATALOG.DATA.GOV/DATASET/ELECTRIC-VEHICLE-POPULATION-DATA



```
# Import Data
data = pd.read_csv('/content/Electric_Vehicle_Population_Data.csv')
data
```

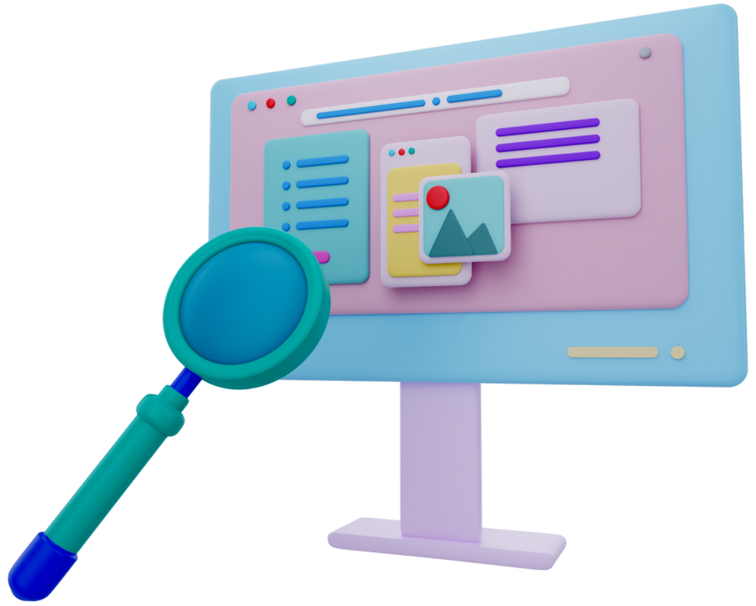
	VIN (1-10)	County	City	State	Postal Code	Model Year	Make	Model	Electric Vehicle Type	Clean Alternative Fuel Vehicle (CAFV) Eligibility	Electric Range	Base MSRP	Legislative District	DOL Vehicle ID	Vehicle Location	Electric Utility	2020 Census Tract
0	1N4AZ0CP5D	Kitsap	Bremerton	WA	98310.0	2013	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	75	0	23.0	214384901	POINT (-122.61136499999998 47.575195000000065)	PUGET SOUND ENERGY INC	5.303508e+10
1	1N4AZ1CP8K	Kitsap	Port Orchard	WA	98366.0	2019	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	150	0	26.0	271008636	POINT (-122.63926499999997 47.53730000000007)	PUGET SOUND ENERGY INC	5.303509e+10
2	5YJXCAE28L	King	Seattle	WA	98199.0	2020	TESLA	MODEL X	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	293	0	36.0	8781552	POINT (-122.394185 47.63919500000003)	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)	5.303301e+10
3	SADHC2S1XK	Thurston	Olympia	WA	98503.0	2019	JAGUAR	I-PACE	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	234	0	2.0	8308492	POINT (-122.8285 47.03646)	PUGET SOUND ENERGY INC	5.306701e+10
4	JN1AZ0CP9B	Snohomish	Everett	WA	98204.0	2011	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	73	0	21.0	245524527	POINT (-122.24128499999995 47.91088000000008)	PUGET SOUND ENERGY INC	5.306104e+10
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
138774	KNAGV4LD6L	Clark	Vancouver	WA	98686.0	2020	KIA	OPTIMA	Plug-in Hybrid Electric Vehicle (PHEV)	Not eligible due to low battery range	28	0	17.0	122272881	POINT (-122.64839529999995 45.70104270000007)	BONNEVILLE POWER ADMINISTRATION||PUD NO 1 OF C...	5.301104e+10
138775	7SAYGDEE1N	Thurston	Lacey	WA	98503.0	2022	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	22.0	187390038	POINT (-122.8285 47.03646)	PUGET SOUND ENERGY INC	5.306701e+10
138776	5YJYGDEE7M	Benton	Richland	WA	99352.0	2021	TESLA	MODEL Y	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	8.0	181300416	POINT (-119.29441499999996 46.27187500000008)	BONNEVILLE POWER ADMINISTRATION||CITY OF RICHL...	5.300501e+10
138777	WBY33AW08P	Pierce	Tacoma	WA	98443.0	2023	BMW	I4	Battery Electric Vehicle (BEV)	Eligibility unknown as battery range has not b...	0	0	25.0	224612055	POINT (-122.36463499999996 47.19449000000003)	BONNEVILLE POWER ADMINISTRATION||CITY OF TACOM...	5.305394e+10
138778	JN1AZ0CP2B	King	Seattle	WA	98118.0	2011	NISSAN	LEAF	Battery Electric Vehicle (BEV)	Clean Alternative Fuel Vehicle Eligible	73	0	37.0	246621546	POINT (-122.28338999999994 47.549285000000054)	CITY OF SEATTLE - (WA)|CITY OF TACOMA - (WA)	5.303301e+10

138779 rows × 17 columns

138779 rows x 17 columns

Deskripsi Dataset

FITUR-FITUR PADA DATASET

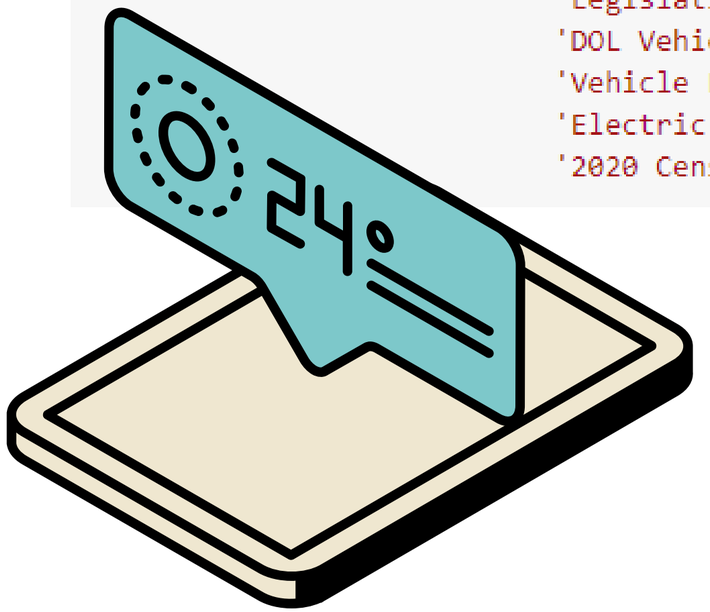


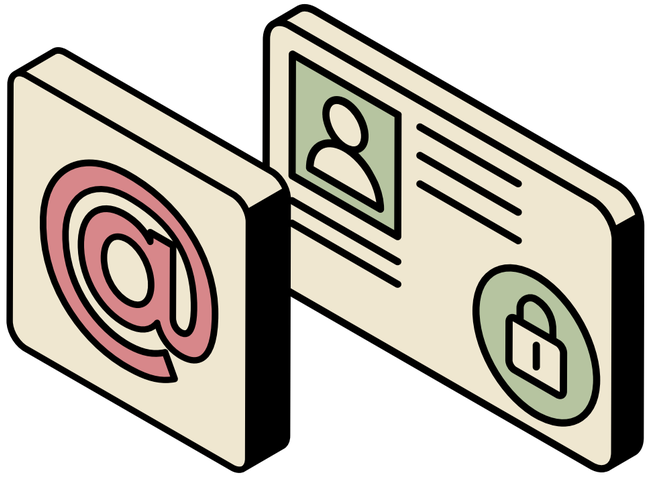
1. **VIN (1-10)**: Vehicle Identification Number
2. **County**: Daerah
3. **City**: Kota
4. **State**: Negara
5. **Postal Code**: Kode Pos
6. **Model Year**: Tahun Pembuatan Modelnya
7. **Make**: Manufacture
8. **Model**: Varian
9. **Electric Vehicle Type**: Jenis Kendaraan Listrik
10. **Clean Alternative Fuel Vehicle (CAFV) Eligibility**: Memenuhi Syarat Sebagai Kendaraan Listrik
11. **Electric Range**: Listrik pada Kendaraan
12. **Base MSRP**: Base Manufacturer Suggested Retail Price
13. **Legislative District**: Kawasan yang Menggunakan Kendaraan Listrik
14. **DOL Vehicle ID**: Department of Licensing
15. **Vehicle Location**: Lokasi Kendaraan
16. **Electric Utility**: Electric Privider/Supplier
17. **2020 Census Tract**: Sensus Daerah 2020

Rename Columns

Mengubah beberapa nama variabel untuk mempermudah.

```
data.rename(columns = {'VIN (1-10)': 'VIN',
                       'Postal Code': 'Zipcode',
                       'Model Year': 'Year',
                       'Electric Vehicle Type': 'EVT',
                       'Clean Alternative Fuel Vehicle (CAFV) Eligibility': 'CAFV',
                       'Electric Range': 'Range',
                       'Base MSRP': 'MSRP',
                       'Legislative District': 'District',
                       'DOL Vehicle ID': 'DOL',
                       'Vehicle Location': 'Location',
                       'Electric Utility': 'Provider',
                       '2020 Census Tract': 'Census2020'}, inplace = True)
```

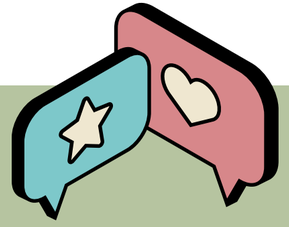




```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 138779 entries, 0 to 138778
Data columns (total 17 columns):
 #   Column      Non-Null Count   Dtype
---  ------      --------------   -----
 0   VIN         138779 non-null  object
 1   County      138776 non-null  object
 2   City        138776 non-null  object
 3   State       138779 non-null  object
 4   Zipcode     138776 non-null  float64
 5   Year        138779 non-null  int64
 6   Make        138779 non-null  object
 7   Model       138493 non-null  object
 8   EVT         138779 non-null  object
 9   CAFV        138779 non-null  object
 10  Range       138779 non-null  int64
 11  MSRP        138779 non-null  int64
 12  District    138464 non-null  float64
 13  DOL         138779 non-null  int64
 14  Location    138773 non-null  object
 15  Provider    138776 non-null  object
 16  Census2020  138776 non-null  float64
dtypes: float64(3), int64(4), object(10)
memory usage: 18.0+ MB
```

Values Observation

Data Cleaning



Melihat jumlah missing values pada dataset

```
data.isna().sum()
```

```
VIN                0
County             3
City               3
State              0
Zipcode            3
Year               0
Make               0
Model            286
EVT                0
CAFV               0
Range              0
MSRP               0
District         315
DOL                0
Location           6
Provider           3
Census2020         3
dtype: int64
```

Melihat jumlah data yang duplikat pada dataset

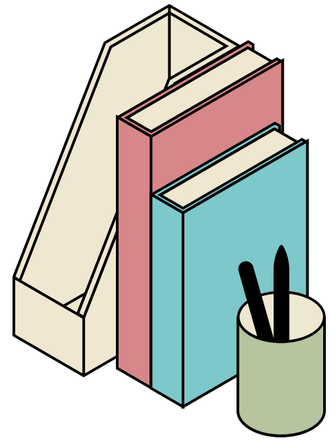
```
data.duplicated().sum()
```

```
0
```

Handling Missing Values

```
# Handling missing value
data = data.dropna()
```

Data Description



Melihat statistik deskriptif dari tiap variabel numerik pada dataset.

	Zipcode	Year	Range	MSRP	District	DOL	Census2020
count	138175.000000	138175.000000	138175.000000	138175.000000	138175.000000	1.381750e+05	1.381750e+05
mean	98258.819685	2019.766441	72.448981	1407.425982	29.427190	2.079269e+08	5.303963e+10
std	302.371730	3.012774	97.809794	9551.756736	14.799921	8.464173e+07	1.617000e+07
min	98001.000000	1997.000000	0.000000	0.000000	1.000000	4.385000e+03	5.300195e+10
25%	98052.000000	2018.000000	0.000000	0.000000	18.000000	1.643437e+08	5.303301e+10
50%	98122.000000	2021.000000	21.000000	0.000000	33.000000	2.085477e+08	5.303303e+10
75%	98370.000000	2022.000000	149.000000	0.000000	43.000000	2.342263e+08	5.305307e+10
max	99403.000000	2024.000000	337.000000	845000.000000	49.000000	4.792548e+08	5.307794e+10

Exploratory Data Analysis

Top Three

Akan dilakukan eksplorasi data untuk top three variabel 'Make' perusahaan pembuat kendaraan listrik terbanyak.

TESLA

63032

NISSAN

13194

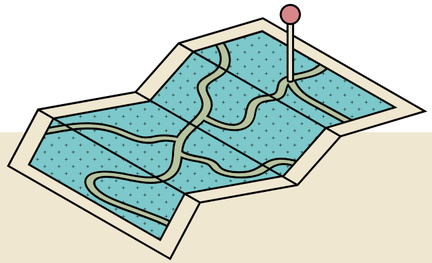
CHEVROLET

11527

```
data.Make.value_counts()
```

```
TESLA                 63032
NISSAN                13194
CHEVROLET             11527
FORD                   7046
BMW                    6014
KIA                    5663
TOYOTA                 4923
VOLKSWAGEN             3704
VOLVO                  3036
JEEP                   2938
AUDI                   2787
HYUNDAI                2625
CHRYSLER               2381
RIVIAN                 1932
PORSCHE                 980
MERCEDES-BENZ           862
FIAT                    818
HONDA                   799
MITSUBISHI              763
MINI                    758
POLESTAR                727
SUBARU                  368
SMART                   276
JAGUAR                  222
LINCOLN                 220
LUCID                   160
CADILLAC                126
LEXUS                    88
GENESIS                  77
MAZDA                    53
LAND ROVER               43
FISKER                   15
AZURE DYNAMICS            8
TH!NK                     4
BENTLEY                   3
WHEEGO ELECTRIC CARS      3
Name: Make, dtype: int64
```

Pivot Table



Membuat Pivot Table data perusahaan top 3 dengan 'Make' sebagai indeks atas, 'Year' sebagai kolom, dan 'Count' sebagai nilai.

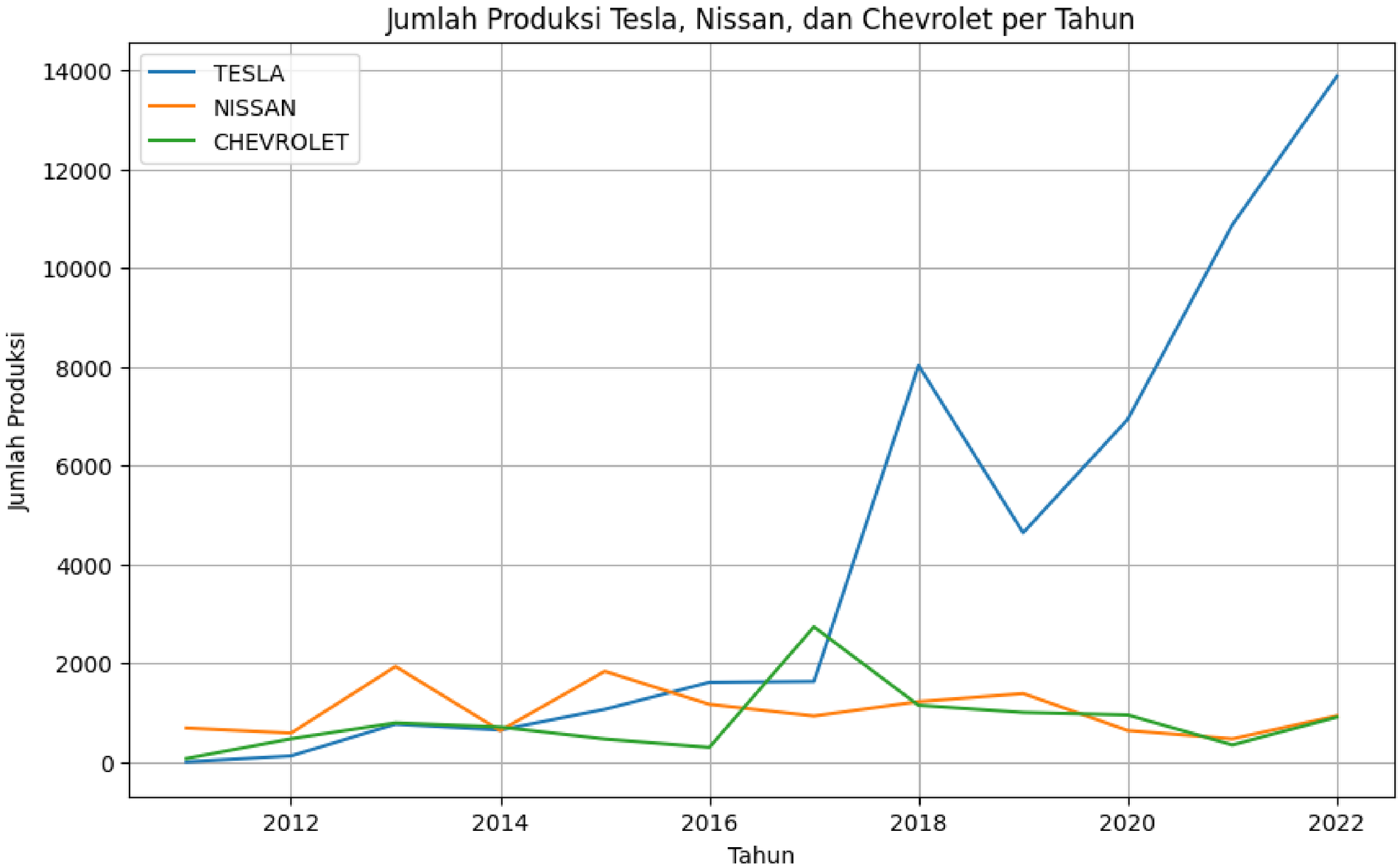
Year	1997	2008	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
Make																
CHEVROLET	1	0	0	76	474	798	713	470	301	2738	1152	1007	957	351	914	1575
NISSAN	0	0	0	694	589	1937	640	1840	1172	936	1226	1391	642	473	940	714
TESLA	0	18	21	7	128	771	653	1071	1617	1635	8027	4642	6938	10872	13879	12753

Drop Row

Dropping baris data tahun 1997, 2008, 2010, dan 2023.

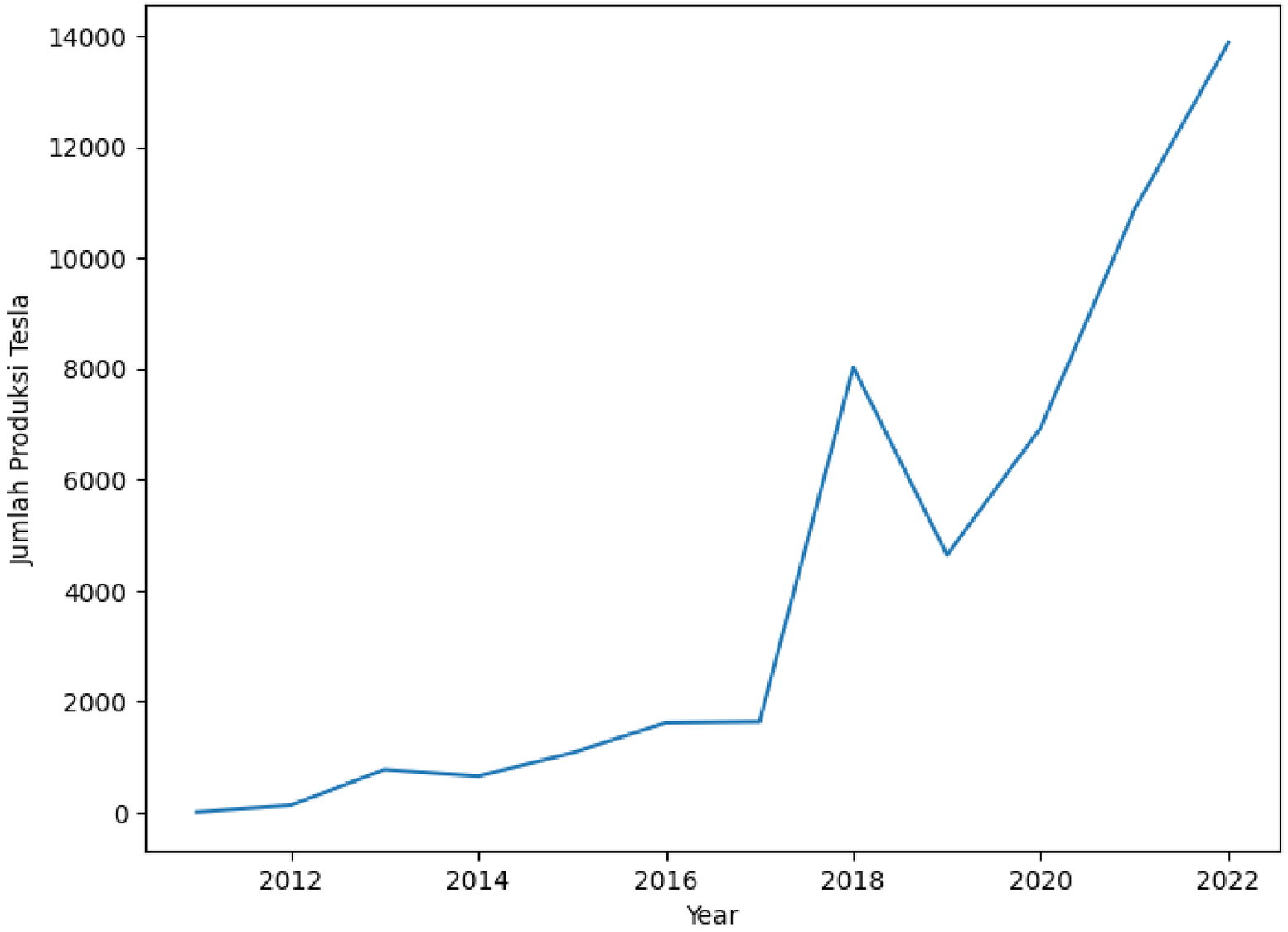
	Year	Make	Count
10	2011	CHEVROLET	76
11	2011	NISSAN	694
12	2011	TESLA	7
15	2012	CHEVROLET	474
19	2012	NISSAN	589
20	2012	TESLA	128
22	2013	CHEVROLET	798
25	2013	NISSAN	1937
27	2013	TESLA	771
31	2014	CHEVROLET	713
37	2014	NISSAN	640
40	2014	TESLA	653
43	2015	CHEVROLET	470
48	2015	NISSAN	1840
51	2015	TESLA	1071
57	2016	CHEVROLET	301
64	2016	NISSAN	1172
67	2016	TESLA	1617
73	2017	CHEVROLET	2738
81	2017	NISSAN	936
84	2017	TESLA	1635
91	2018	CHEVROLET	1152
101	2018	NISSAN	1226
104	2018	TESLA	8027
110	2019	CHEVROLET	1007
121	2019	NISSAN	1391
125	2019	TESLA	4642
132	2020	CHEVROLET	957
144	2020	NISSAN	642
146	2020	TESLA	6938
152	2021	CHEVROLET	351
164	2021	NISSAN	473
167	2021	TESLA	10872
174	2022	CHEVROLET	914
187	2022	NISSAN	940
191	2022	TESLA	13879

Data Visualization: Plot Perbandingan Jumlah Produksi Mobil Top 3 per Tahun



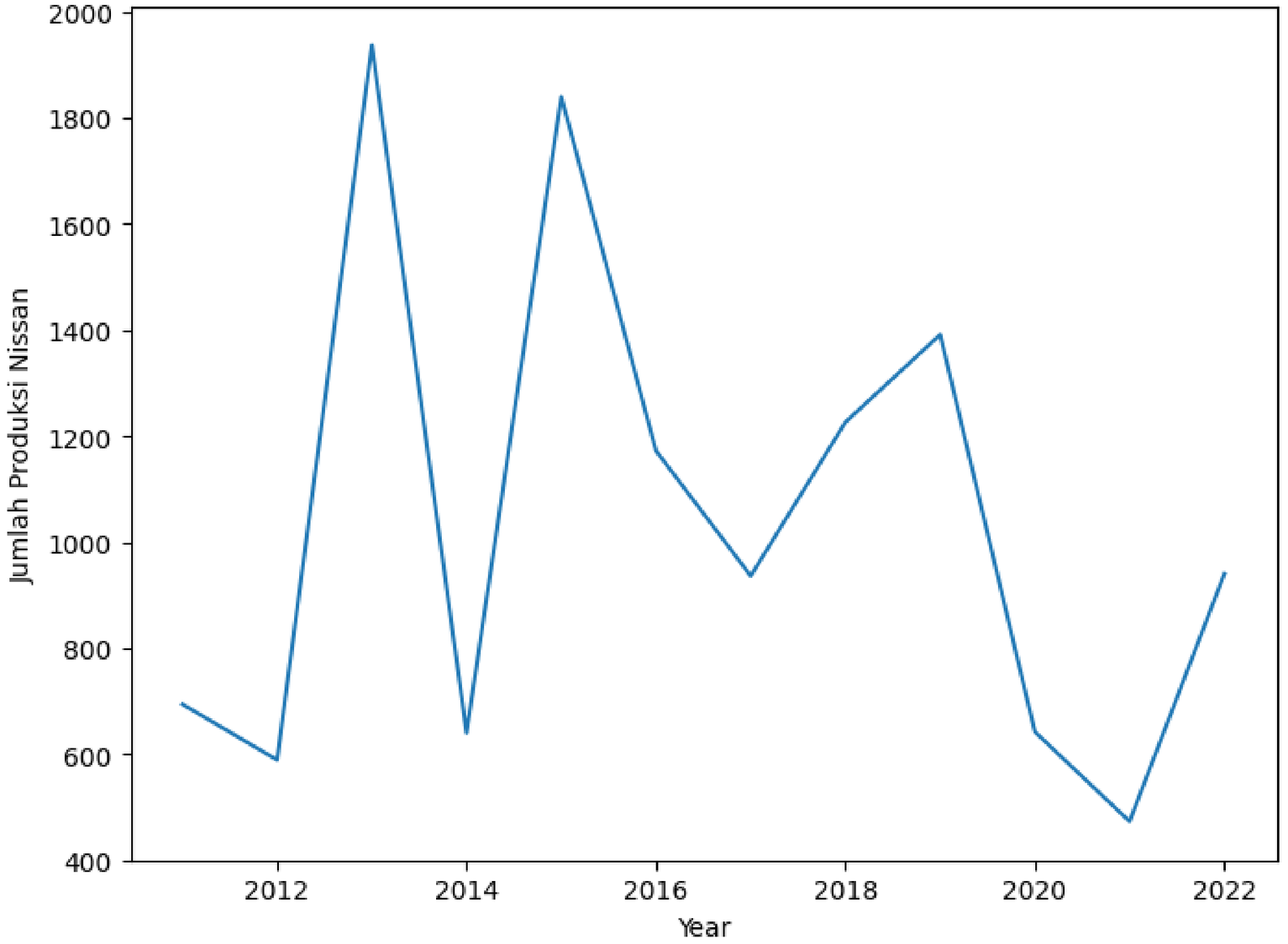
Data Visualization: Jumlah Produksi TESLA per Tahun

	Year	Make	Count
12	2011	TESLA	7
20	2012	TESLA	128
27	2013	TESLA	771
40	2014	TESLA	653
51	2015	TESLA	1071
67	2016	TESLA	1617
84	2017	TESLA	1635
104	2018	TESLA	8027
125	2019	TESLA	4642
146	2020	TESLA	6938
167	2021	TESLA	10872
191	2022	TESLA	13879



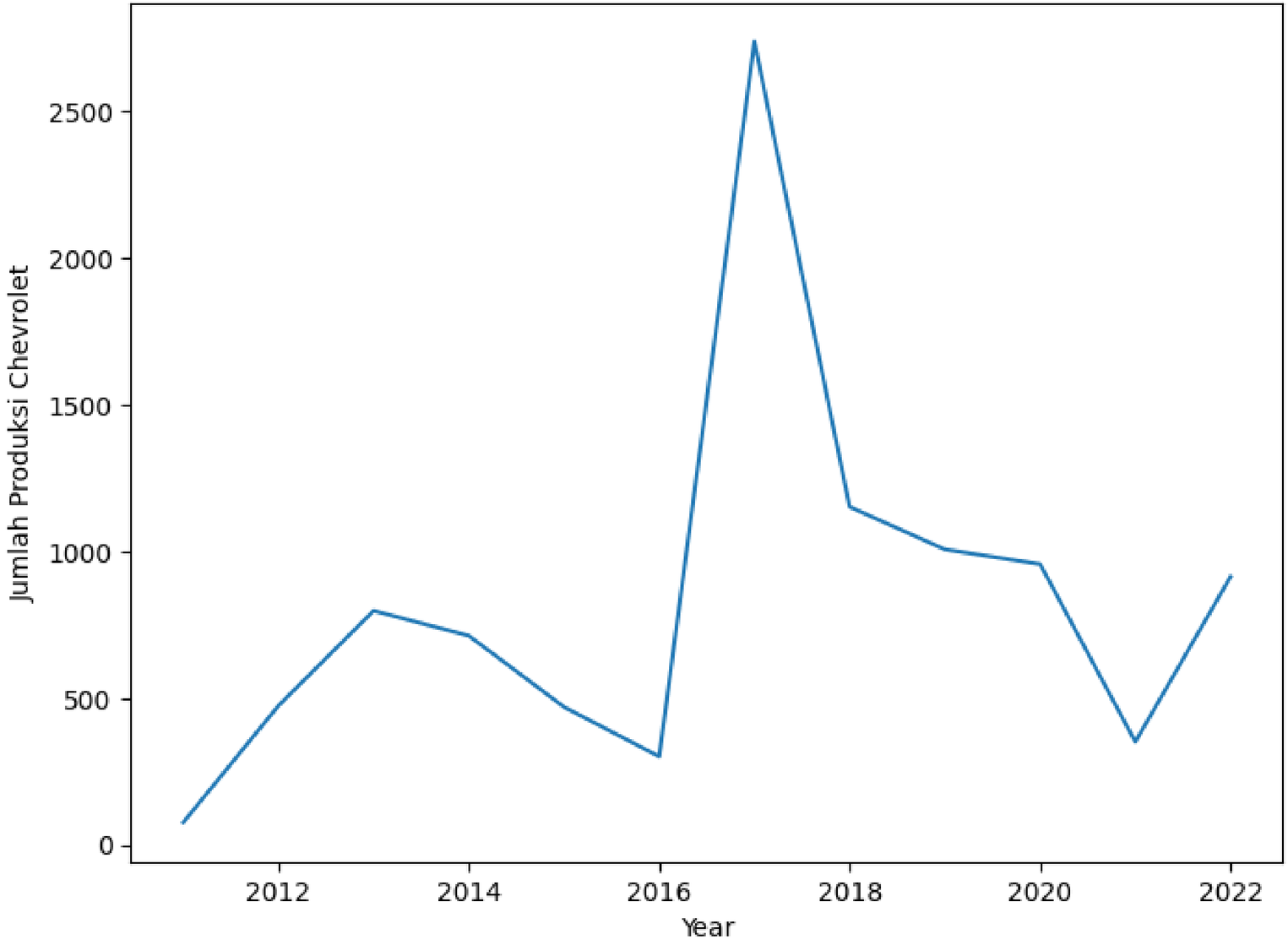
Data Visualization: Jumlah Produksi NISSAN per Tahun

	Year	Make	Count
11	2011	NISSAN	694
19	2012	NISSAN	589
25	2013	NISSAN	1937
37	2014	NISSAN	640
48	2015	NISSAN	1840
64	2016	NISSAN	1172
81	2017	NISSAN	936
101	2018	NISSAN	1226
121	2019	NISSAN	1391
144	2020	NISSAN	642
164	2021	NISSAN	473
187	2022	NISSAN	940



Data Visualization: Jumlah Produksi CHEVROLET per Tahun

	Year	Make	Count
10	2011	CHEVROLET	76
15	2012	CHEVROLET	474
22	2013	CHEVROLET	798
31	2014	CHEVROLET	713
43	2015	CHEVROLET	470
57	2016	CHEVROLET	301
73	2017	CHEVROLET	2738
91	2018	CHEVROLET	1152
110	2019	CHEVROLET	1007
132	2020	CHEVROLET	957
152	2021	CHEVROLET	351
174	2022	CHEVROLET	914



Feature Engineering

Feature Engineering

Membuat dataframe baru dengan dropping fitur-fitur selain 'Year' dan 'Make' lalu menambahkan kolom 'Count'.

```
# Jumlah pembelian mobil setiap tahunnya
data = data.groupby(['Year', 'Make']).size().reset_index(name='Count')
```

Spesifikasi dataframe baru tersebut dengan hanya menggunakan data row terkait perusahaan top three yang ingin diteliti lebih lanjut.

```
# Keep only Tesla, Nissan, and Chevrolet
data = data[data['Make'].isin(['TESLA', 'NISSAN', 'CHEVROLET'])]
```

Modeling & Evaluation

Splitting Data

Splitting Features & Target

X : Year
y : Count

Splitting Training & Testing Set

Train Data
Test Data

Year	Count
2011	
2012	
2013	
2014	
2015	
2016	
2017	
2018	
2019	
2020	
2021	
2022	

Hyperparameter Tuning (tune the degree)

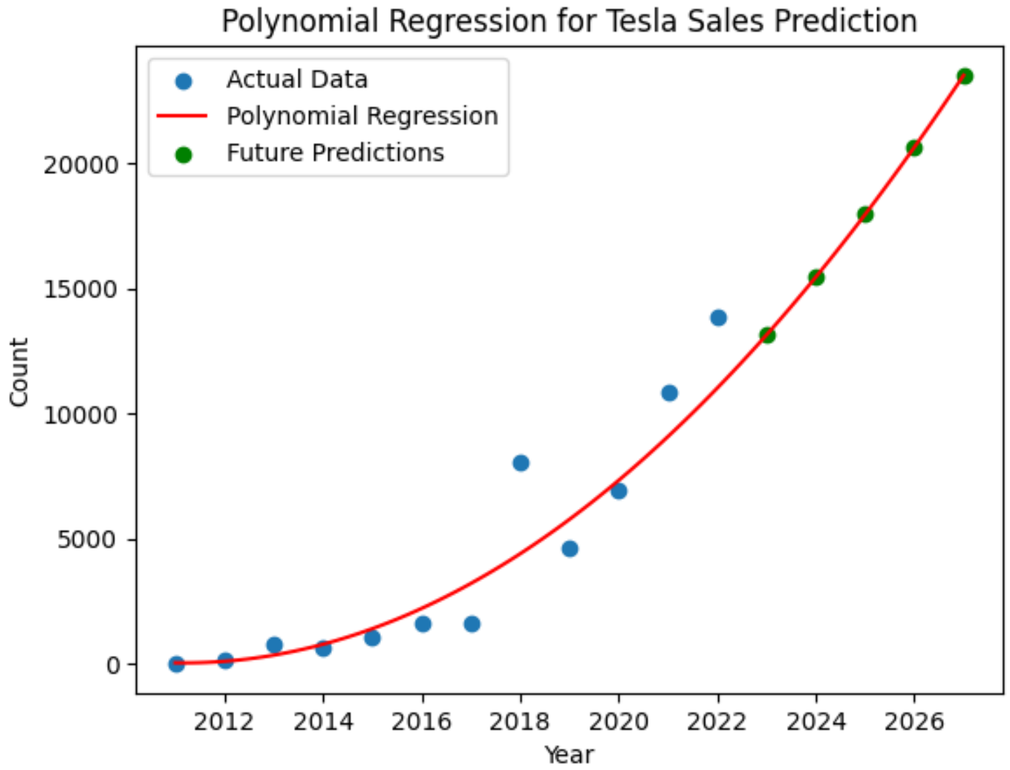
Digunakan Metrik Evaluasi MAE (Mean Absolute Error).

	TESLA		NISSAN		CHEVROLET	
Degree	Test Data	Train Data	Test Data	Train Data	Test Data	Train Data
6	50062.38182	793.7313287	8102.018182	245.851655	31671.41818	291.3435431
5	11163.99394	827.6626573	7116.224242	246.7995338	5392.824242	374.1683916
4	11044.39394	828.8657343	1481.590909	331.1714452	3030.742424	380.4413753
3	2715.939394	616.4179487	315.2348485	348.9004662	844.1651515	404.469697
2	2333.121212	823.28	408.9893939	338.5939394	364.4287879	370.9975758

Model Evaluation

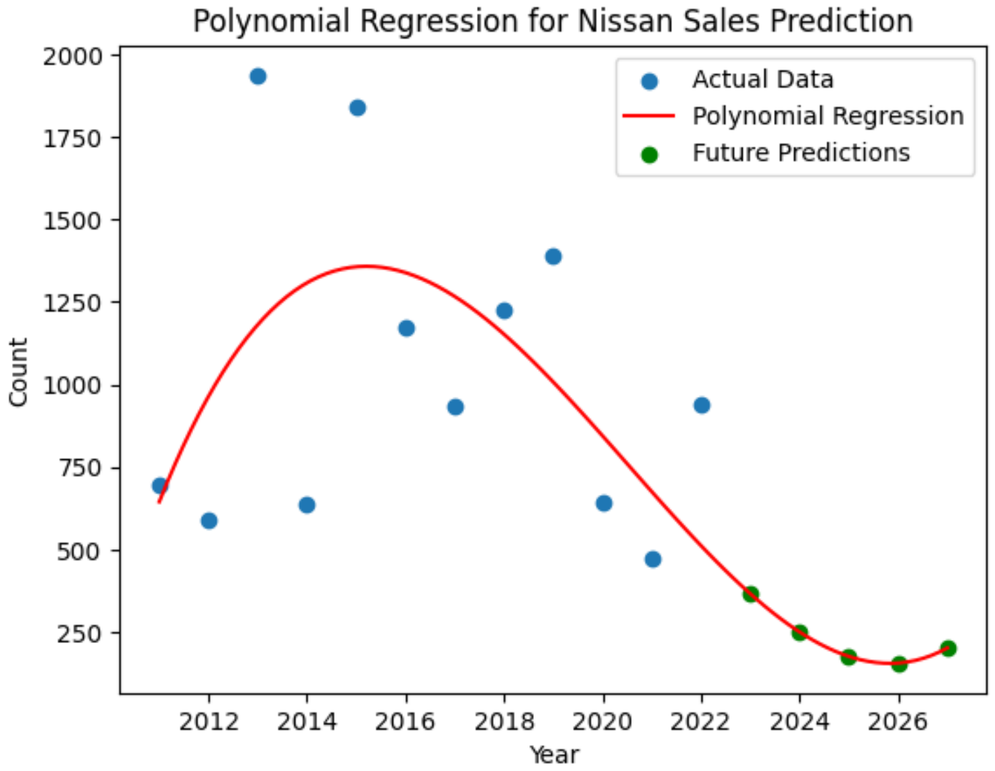
	TESLA		NISSAN		CHEVROLET	
Matrix	Test Data	Train Data	Test Data	Train Data	Test Data	Train Data
r2_score	-1.534518481	0.76828912	-1.056265766	0.225664826	-1.656802688	0.29139985
MAE	2333.121212	823.28	315.2348485	348.9004662	364.4287879	370.9975758
MSE	5729310.073	1775370.217	112112.2362	172496.951	210531.0228	346938.598

Tesla Prediction



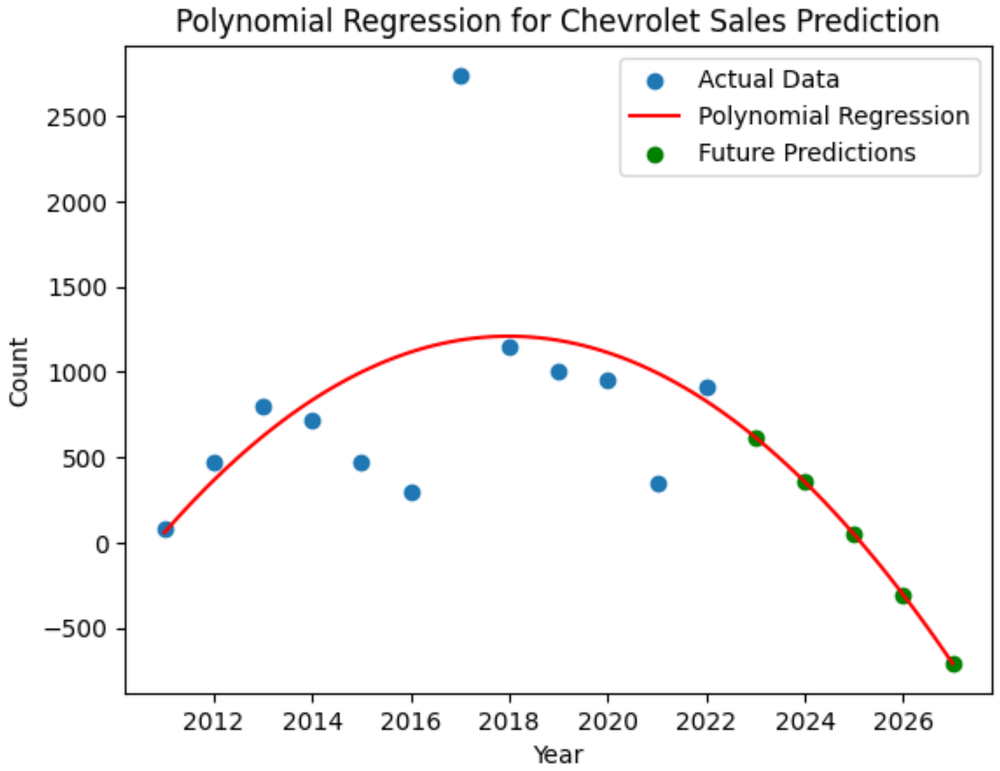
Year: 2023, Predicted Count: 13136.76
Year: 2024, Predicted Count: 15450.15
Year: 2025, Predicted Count: 17951.39
Year: 2026, Predicted Count: 20640.47
Year: 2027, Predicted Count: 23517.41

Nissan Prediction



Year: 2023, Predicted Count: 366.23
Year: 2024, Predicted Count: 250.39
Year: 2025, Predicted Count: 176.39
Year: 2026, Predicted Count: 156.25
Year: 2027, Predicted Count: 201.99

Chevrolet Prediction



Year: 2023, Predicted Count: 615.33
Year: 2024, Predicted Count: 355.12
Year: 2025, Predicted Count: 47.74
Year: 2026, Predicted Count: -306.81
Year: 2027, Predicted Count: -708.54

Konklusi & Rekomendasi

Konklusi

1. Prediksi menunjukkan potensi pertumbuhan yang besar khususnya bagi Tesla sehingga hal ini menandakan peluang yang menarik dalam industri mobil listrik.
2. Informasi ini dapat membantu merek mengarahkan strategi pemasaran dengan lebih tepat. Kampanye dapat diarahkan ke area pertumbuhan potensial.
3. Untuk Nissan dan Chevrolet, tren penjualan dapat dipahami sehingga membantu perencanaan produksi yang lebih efisien, menghindari biaya yang tidak perlu.

Rekomendasi

1. Perlu eksplorasi model yang lebih sederhana dan memiliki kemampuan generalisasi yang lebih baik untuk meningkatkan akurasi.
2. Integrasi data lebih lengkap, termasuk faktor eksternal, memperkaya analisis.
3. Jika hasil analisis ini digunakan dalam pengambilan keputusan bisnis, penting untuk memantau kinerja prediksi secara teratur dan memperbarui model sesuai dengan kondisi pasar yang berkembang.

Thank you.