Analisis Perkembangan EV di Amerika

Calculus Clan COMPFEST 2023

Problem Background/Description

Kendaraan listrik (EV) telah muncul sebagai solusi potensial yang kuat untuk mengatasi sejumlah tantangan lingkungan, energi, dan transportasi yang dihadapi oleh Amerika Serikat. Sebagai negara dengan salah satu tingkat emisi gas rumah kaca tertinggi di dunia dan ketergantungan pada bahan bakar fosil yang signifikan, AS telah mengidentifikasi adopsi kendaraan listrik sebagai langkah penting dalam rangka mengurangi dampak lingkungan dan memitigasi risiko perubahan iklim. Namun, di balik potensi positifnya, integrasi EV di AS juga dihadapkan pada serangkaian tantangan multidimensi yang memerlukan perhatian serius.

Kami ingin mencoba melakukan analisis perkembangan EV di Amerika Serikat dan juga mencoba membuatkan model prediksinya. Dari hasil analisis dan prediksi kami tersebut, diharapkan dapat membantu produsen mobil EV dan pemerintah dalam peningkatan pertumbuhan EV dengan memberikan gambaran pertumbuhan penjualan EV di masa depan.

Our Objectives

Goals:

1. Signifikansi dampak insentif pajak terhadap perkembangan penggunaan Electric Vehicle di Amerika Serikat
2. Membuat prediksi gambaran pertumbuhan penjualan EV di masa depan

01

Data Preprocessing

[Dataset Electric Vehicle Title and Registration Activity]

Description: Menunjukkan **Records of Title Activity** (transaksi yang mencatat perubahan kepemilikan), dan **Registration Activity** (transaksi yang mengizinkan kendaraan untuk digunakan di jalan umum Washington).

Dataset Possible Issues

Pada dataset yang digunakan terdapat:

- Sale Price yang tidak terecord
- Attribute 2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility, Meets 2019 HB 2042 Electric Range Requirement, Meets 2019 HB 2042 Sale Date Requirement, Meets 2019 HB 2042 Sale Price/Value Requirement, 2019 HB 2042: Battery Range Requirement, 2019 HB 2042: Purchase Date Requirement, 2019 HB 2042: Sale Price/Value Requirement tidak terecord secara lengkap dan tidak konsisten sehingga banyak data yang berisikan Sale Price = 0.

```
data_3.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 761415 entries, 0 to 761414
Data columns (total 35 columns):
 #   Column                                                 Non-Null Count   Dtype
---  ------                                                 --------------   -----
 0   Clean Alternative Fuel Vehicle Type                    761415 non-null  object
 1   VIN (1-10)                                             761415 non-null  object
 2   DOL Vehicle ID                                         761415 non-null  int64
 3   Model Year                                             761415 non-null  int64
 4   Make                                                   761415 non-null  object
 5   Model                                                  760843 non-null  object
 6   Vehicle Primary Use                                    761415 non-null  object
 7   Electric Range                                         761415 non-null  int64
 8   Odometer Reading                                       761415 non-null  int64
 9   Odometer Code                                          761415 non-null  object
 10  New or Used Vehicle                                    761415 non-null  object
 11  Sale Price                                             761415 non-null  int64
 12  Sale Date                                              227956 non-null  object
 13  Base MSRP                                              761415 non-null  int64
 14  Transaction Type                                       761415 non-null  object
 15  DOL Transaction Date                                   761415 non-null  datetime64[ns]
 16  Transaction Year                                       761415 non-null  int64
 17  County                                                 761380 non-null  object
 18  City                                                   761345 non-null  object
 19  State of Residence                                     761414 non-null  object
 20  Postal Code                                            761370 non-null  float64
 21  2015 HB 2778 Exemption Eligibility                     761415 non-null  object
 22  2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility 761415 non-null  object
 23  Meets 2019 HB 2042 Electric Range Requirement          761415 non-null  bool
 24  Meets 2019 HB 2042 Sale Date Requirement               761415 non-null  bool
 25  Meets 2019 HB 2042 Sale Price/Value Requirement        761415 non-null  bool
 26  2019 HB 2042: Battery Range Requirement                761415 non-null  object
 27  2019 HB 2042: Purchase Date Requirement                761415 non-null  object
 28  2019 HB 2042: Sale Price/Value Requirement             761415 non-null  object
 29  Electric Vehicle Fee Paid                              761415 non-null  object
 30  Transportation Electrification Fee Paid                670384 non-null  object
 31  Hybrid Vehicle Electrification Fee Paid                670384 non-null  object
 32  2020 Census Tract                                      761380 non-null  float64
 33  Legislative District                                   758791 non-null  float64
 34  Electric Utility                                       761380 non-null  object
dtypes: bool(3), datetime64[ns](1), float64(3), int64(7), object(21)
memory usage: 188.1+ MB
```

Data Information

Terdapat 34 feature/kolom
Dengan **total data 761515** baris

- VIN = Vehicle Identification Number
- MSRP = MSRP stands for manufacturer's suggested retail price. The MSRP is the suggested sticker price you see on a car window, and it is the price the manufacturer suggests the dealer ask for the vehicle

Missing Values

Missing values ditemukan pada **11 feature** pada dataset.

Dengan missing values terbanyak ditemukan di feature **Sale Date, Transportation Electrification Fee Paid,** dan **Hybrid Vehicle Electrification Fee Paid**.

Perlu dilakukannya data cleaning dan preprocessing lebih lanjut untuk *handling missing values*.

```
data_3.isnull().sum()
```

```
Clean Alternative Fuel Vehicle Type                                0
VIN (1-10)                                                         0
DOL Vehicle ID                                                     0
Model Year                                                         0
Make                                                               0
Model                                                            572
Vehicle Primary Use                                                0
Electric Range                                                     0
Odometer Reading                                                   0
Odometer Code                                                      0
New or Used Vehicle                                                0
Sale Price                                                         0
Sale Date                                                     533459
Base MSRP                                                          0
Transaction Type                                                   0
DOL Transaction Date                                               0
Transaction Year                                                   0
County                                                            35
City                                                              70
State of Residence                                                 1
Postal Code                                                       45
2015 HB 2778 Exemption Eligibility                                 0
2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility     0
Meets 2019 HB 2042 Electric Range Requirement                      0
Meets 2019 HB 2042 Sale Date Requirement                           0
Meets 2019 HB 2042 Sale Price/Value Requirement                    0
2019 HB 2042: Battery Range Requirement                            0
2019 HB 2042: Purchase Date Requirement                            0
2019 HB 2042: Sale Price/Value Requirement                         0
Electric Vehicle Fee Paid                                          0
Transportation Electrification Fee Paid                        91031
Hybrid Vehicle Electrification Fee Paid                        91031
2020 Census Tract                                                 35
Legislative District                                            2624
Electric Utility                                                  35
dtype: int64
```

Handling Missing Values per Feature

```
#Drop NaN values nya karena presentasenya kecil
data_3.dropna(subset = ['Model'], inplace=True)
data_3.reset_index(drop = True, inplace = True)
print("Presentase Missing Values kolom Model: ", (100*(data_3['Model'].isnull().sum()/len(data_3['Model']))))
data_3['Model'].isnull().sum()
```

```
Presentase Missing Values kolom Model:  0.0
0
```

```
data_3 = data_3.fillna('Non-Sale Transaction')
```

Kolom Transportation and Hybrid Elec fee paid*

```
print(set(data_3['Transportation Electrification Fee Paid']) == set(data_3['Hybrid Vehicle Electrification Fee Paid']))
print("Presentase Missing Values kolom TEFP: ", (100*(data_3['Transportation Electrification Fee Paid'].isnull().sum()/len(data_3['Transportation Electrification Fee
```

```
True
Presentase Missing Values kolom TEFP:  11.963647022791593
```

```
print('Value counts TFEP: \n',data_3['Transportation Electrification Fee Paid'].value_counts(dropna=False))
print()
print('Value counts HVEFP: \n', data_3['Hybrid Vehicle Electrification Fee Paid'].value_counts(dropna=False))
#diketahui apa yang hilang di TFEP jg hilang di HVEFP
```

```
Value counts TFEP:
 No               264455
Yes              228084
Not Applicable   177209
NaN               91015
Name: Transportation Electrification Fee Paid, dtype: int64

Value counts HVEFP:
 No               451768
Not Applicable   177209
NaN               91015
Yes               40771
Name: Hybrid Vehicle Electrification Fee Paid, dtype: int64
```

```
# explicitly require this experimental feature
from sklearn.experimental import enable_iterative_imputer # noqa
from sklearn.ensemble import RandomForestRegressor, RandomForestClassifier
# now you can import normally from sklearn.impute
from sklearn.impute import IterativeImputer
from sklearn.preprocessing import LabelEncoder

encoders = dict()

for col_name in y_tes.columns:
    series = y_tes[col_name]
    label_encoder = LabelEncoder()
    y_tes[col_name] = pd.Series(
        label_encoder.fit_transform(series[series.notnull()]),
        index=series[series.notnull()].index
    )
    encoders[col_name] = label_encoder

imp = IterativeImputer(estimator=RandomForestClassifier(),
                       initial_strategy='most_frequent',
                       max_iter=10, random_state=0)
# Fit to the dataset containing missing values
imp.fit(y_tes)
# Transform the dataset containing missing values
df_y_tes = pd.DataFrame(imp.transform(y_tes), columns = y_tes.columns)

df_y_tes['Hybrid Vehicle Electrification Fee Paid'] = df_y_tes['Hybrid Vehicle Electrification Fee Paid'].astype('int')
df_y_tes['Transportation Electrification Fee Paid'] = df_y_tes['Transportation Electrification Fee Paid'].astype('int')
df_y_tes['Hybrid Vehicle Electrification Fee Paid'] = pd.DataFrame(label_encoder.inverse_transform(df_y_tes['Hybrid Vehicle Electrification Fee Paid']))
df_y_tes['Transportation Electrification Fee Paid'] = pd.DataFrame(label_encoder.inverse_transform(df_y_tes['Transportation Electrification Fee Paid']))

#DIlakukan semi supervised machine learning imputation karena data missingnya at random
df_y_tes
```

After handling Missing Values per Feature

```
data_3.isnull().sum()

Clean Alternative Fuel Vehicle Type                                   0
VIN (1-10)                                                            0
DOL Vehicle ID                                                        0
Model Year                                                            0
Make                                                                  0
Model                                                                 0
Vehicle Primary Use                                                   0
Electric Range                                                        0
Odometer Reading                                                      0
Odometer Code                                                         0
New or Used Vehicle                                                   0
Sale Price                                                            0
Sale Date                                                             0
Base MSRP                                                             0
Transaction Type                                                      0
DOL Transaction Date                                                  0
Transaction Year                                                      0
County                                                                0
City                                                                  0
State of Residence                                                    0
Postal Code                                                           0
2015 HB 2778 Exemption Eligibility                                    0
2019 HB 2042 Clean Alternative Fuel Vehicle (CAFV) Eligibility        0
Meets 2019 HB 2042 Electric Range Requirement                         0
Meets 2019 HB 2042 Sale Date Requirement                              0
Meets 2019 HB 2042 Sale Price/Value Requirement                       0
2019 HB 2042: Battery Range Requirement                               0
2019 HB 2042: Purchase Date Requirement                               0
2019 HB 2042: Sale Price/Value Requirement                            0
Electric Vehicle Fee Paid                                             0
Transportation Electrification Fee Paid                               0
Hybrid Vehicle Electrification Fee Paid                               0
2020 Census Tract                                                     0
Electric Utility                                                      0
dtype: int64
```

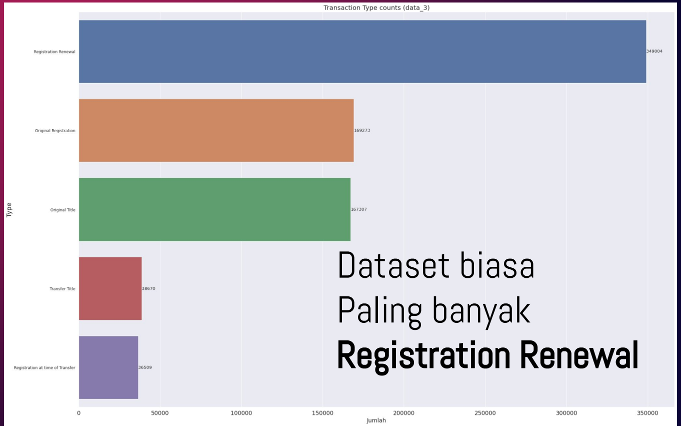
```
print(data_3.shape)

(760763, 34)
```

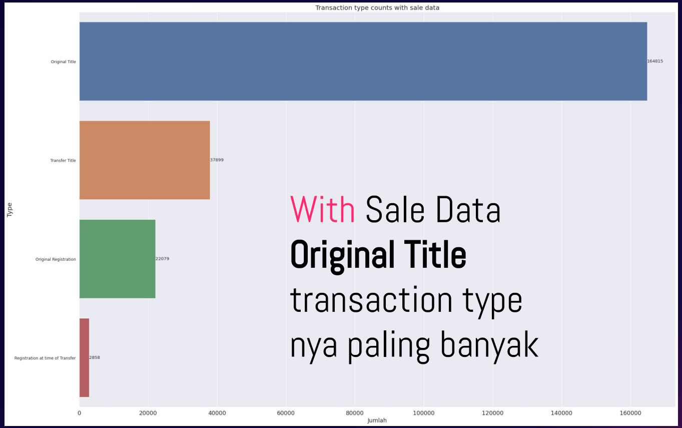
Vehicle titles show proof of vehicle ownership, while vehicle registration signifies a vehicle is registered with the state and cleared for driving on public roads.

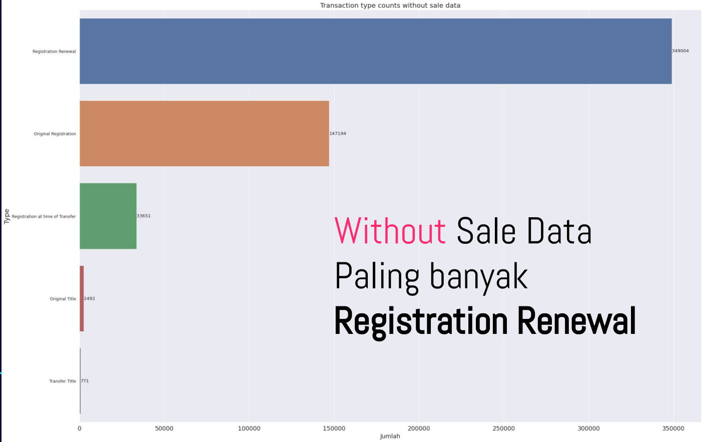
Exploration

Exploratory Data Analysis

Transaction Type

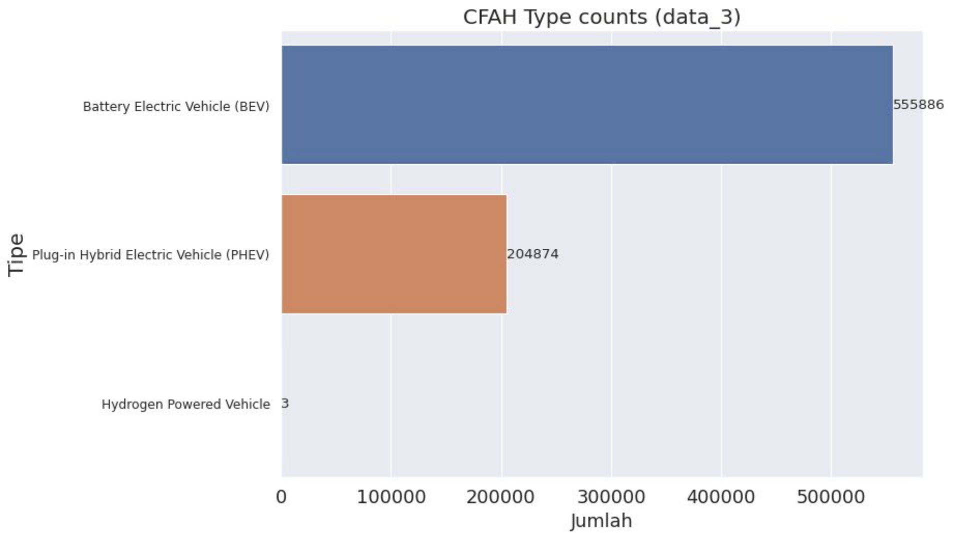






- Dapat dilihat bahwa transaction type data dengan sale date banyak terjadi transaksi jual-beli, dikuatkan dengan status 'original title yang dominan'
- Dan yang tanpa sale date banyak terjadi pembaruan registrasi

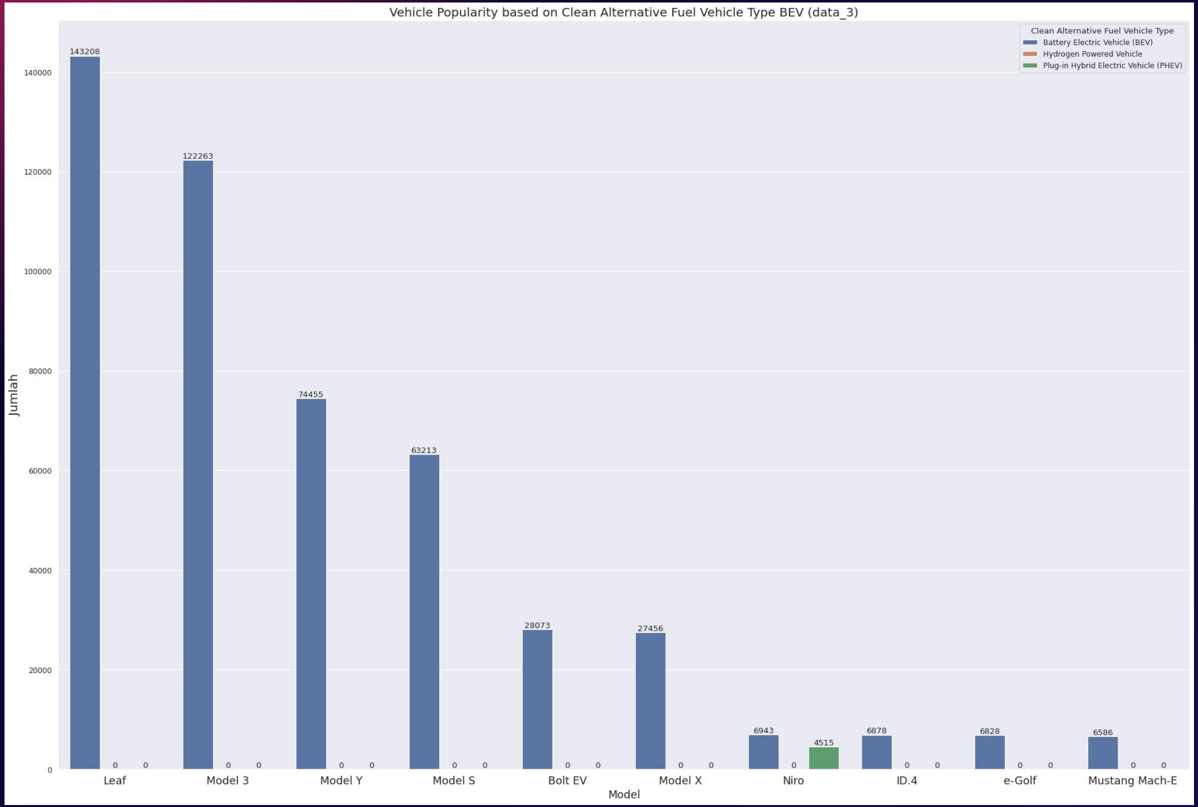
Clean Alternative Fuel Vehicle Type



Tipe Clean Alternative Fuel Vehicle terbanyak adalah Battery Electric Vehicle

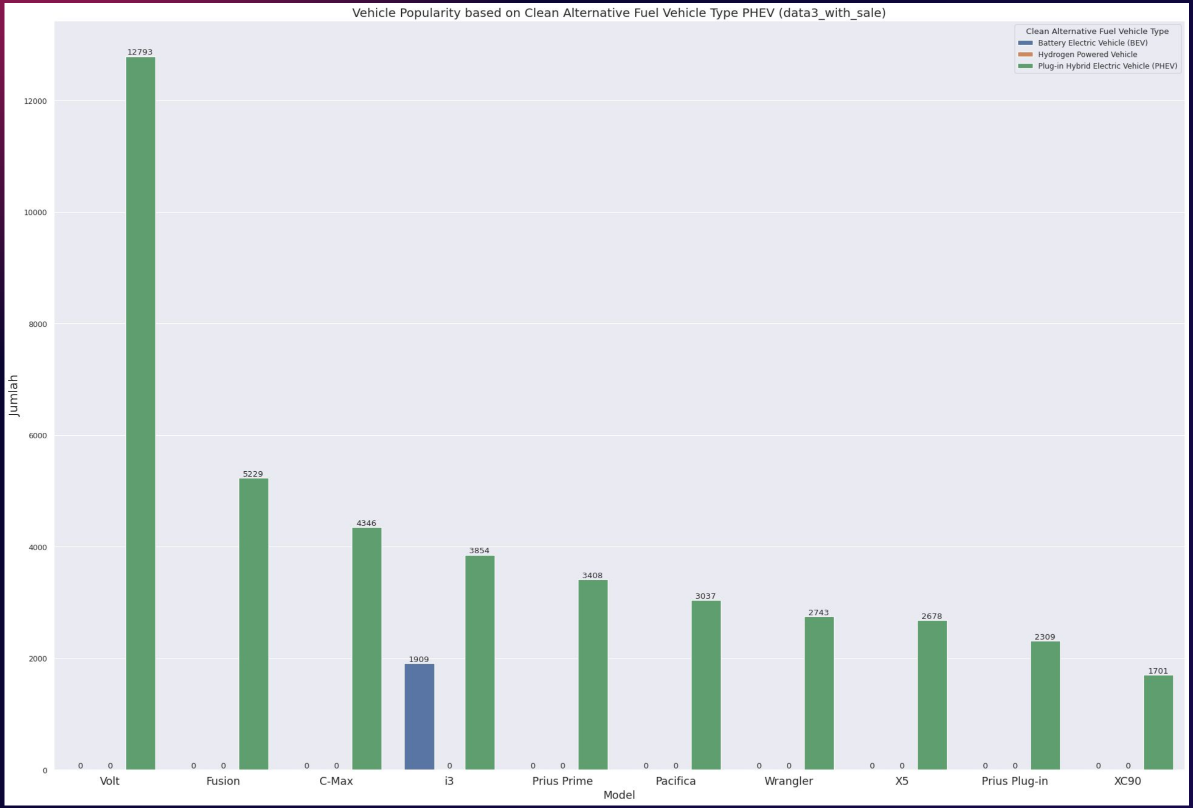
Hal ini dikarenakan tipe BEV memiliki keunggulan yang dipengaruhi faktor peningkatan teknologi baterai (menghasilkan peningkatan kapasitas penyimpanan energi dan jangkauan yang signifikan), selain itu tipe BEV juga diluncurkan dengan variasi desain yang lebih menarik dan kompetitif.

Top 10 Model Mobil CAFV Tipe BEV



Didapat top 10 mobil dengan tipe CFAH BEV paling banyak adalah: ['Leaf', 'Model 3', 'Model Y', 'Model S', 'Bolt EV', 'Model X', 'Niro', 'ID.4', 'e-Golf', 'Mustang Mach-E'].

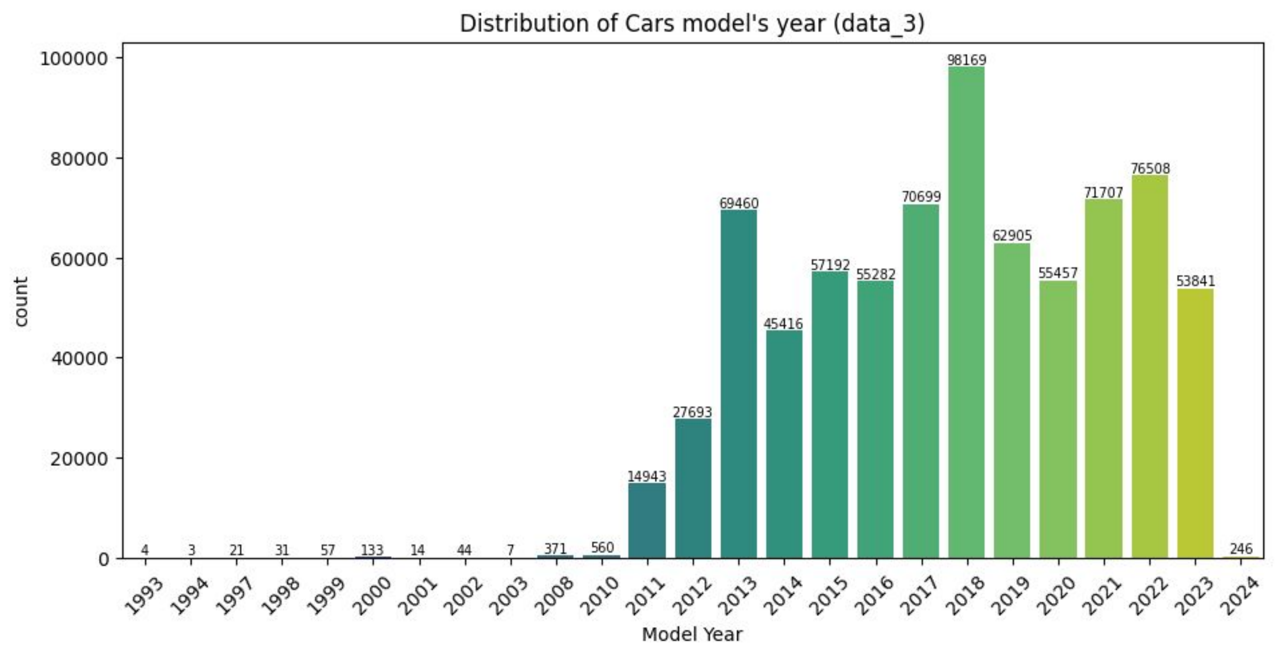
Top 10 Model Mobil CAFV Tipe PHEV



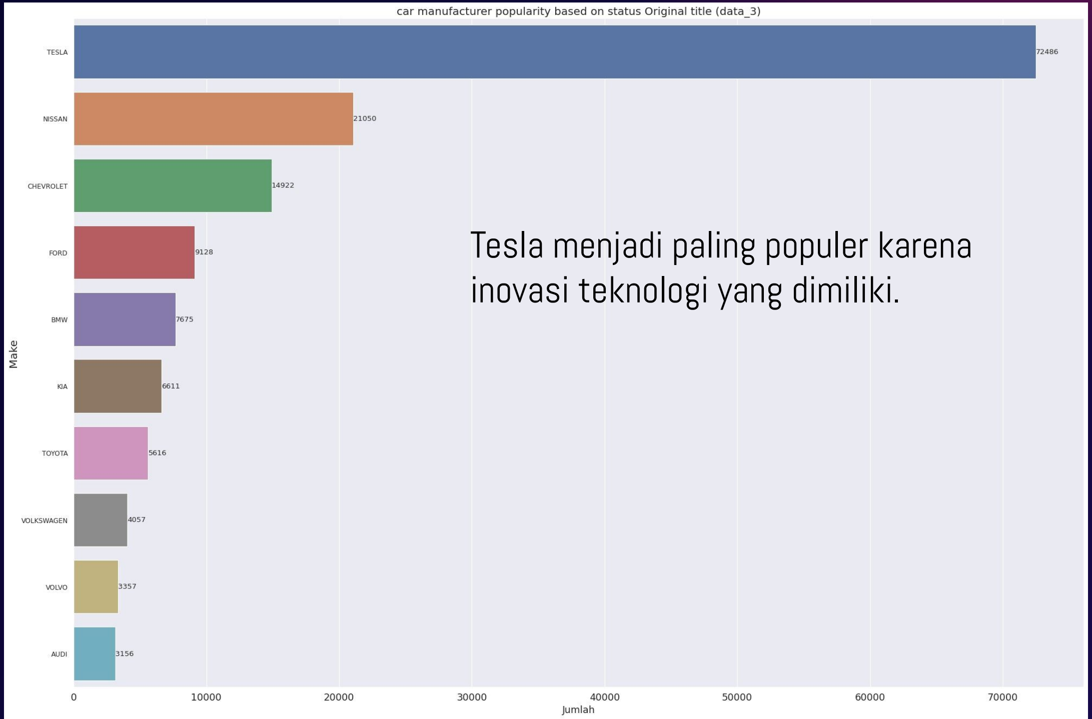
Didapat top 10 mobil dengan tipe CFAH PHEV paling banyak adalah: ['Volt', 'Fusion', 'C-Max', 'i3', 'Prius Prime', 'Pacifica', 'Wrangler', 'X5', 'Prius Plug-in', 'XC90']

Distribution of Cars model's Years

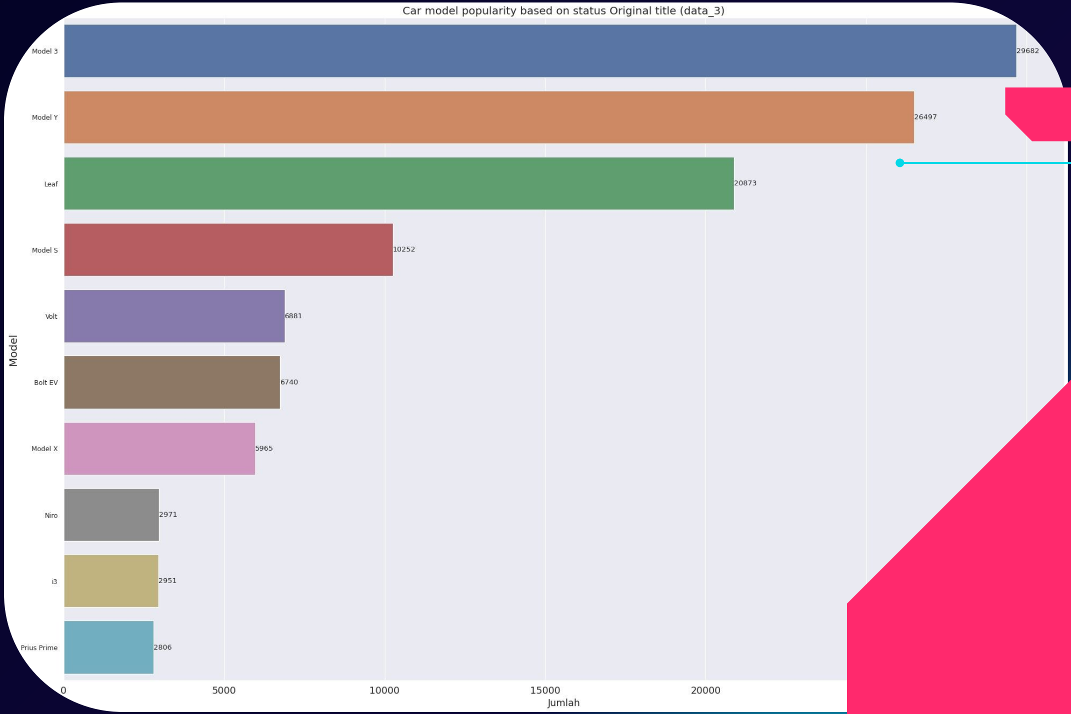
Model EV yang paling banyak digunakan adalah yang diproduksi pada tahun 2018. Disusul dengan produksi dari tahun 2022 dan 2021.



TOP 10 Merek Popularity Based on Original Title: ['TESLA', 'NISSAN', 'CHEVROLET', 'FORD', 'BMW', 'KIA', 'TOYOTA', 'VOLKSWAGEN', 'VOLVO', 'AUDI']

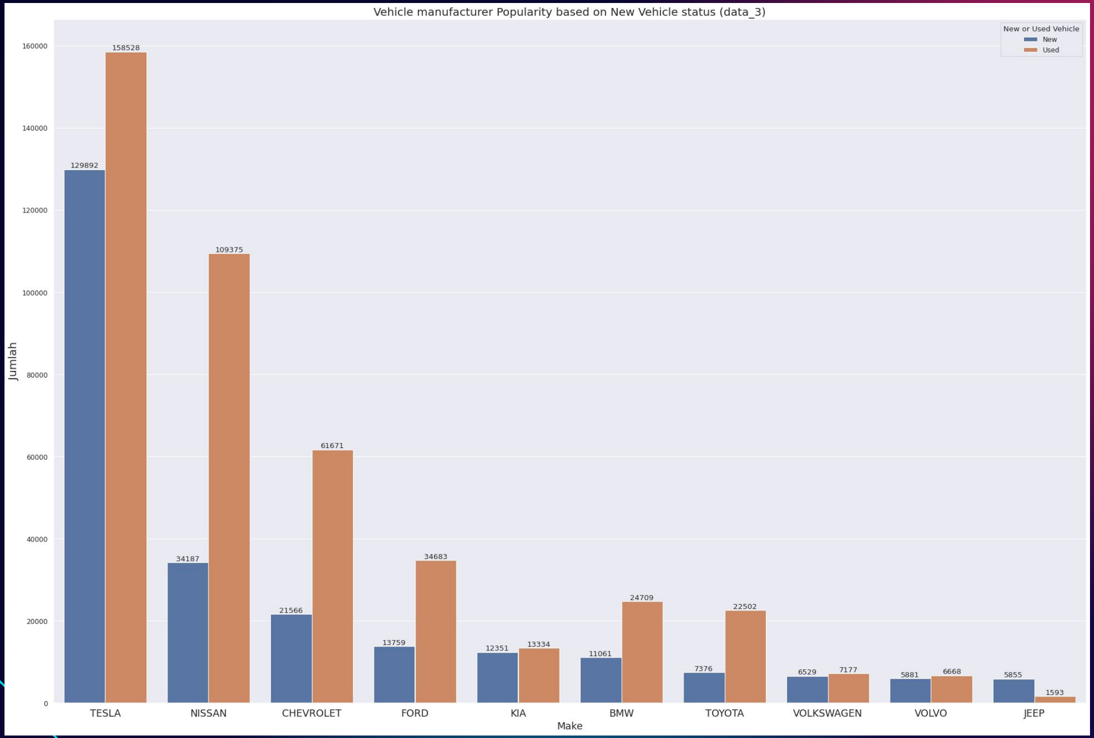


Tesla menjadi paling populer karena inovasi teknologi yang dimiliki.



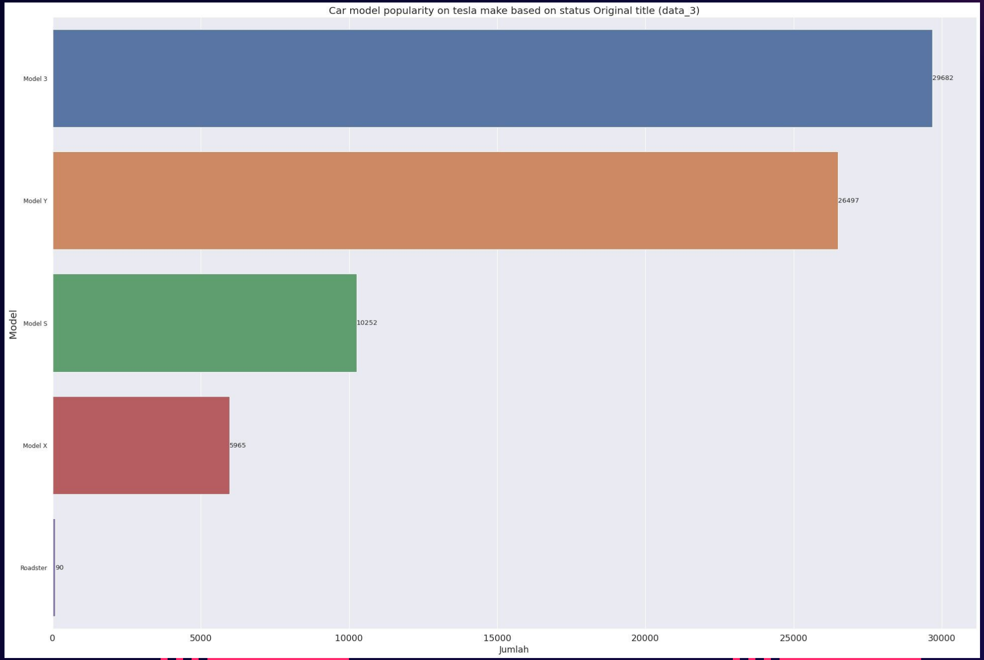
10 Model Terpopuler berdasarkan Original Title

Model 3, Model Y, Leaf, Model S, Volt, Bolt EV, Model X, Niro, i3, Prius Prime



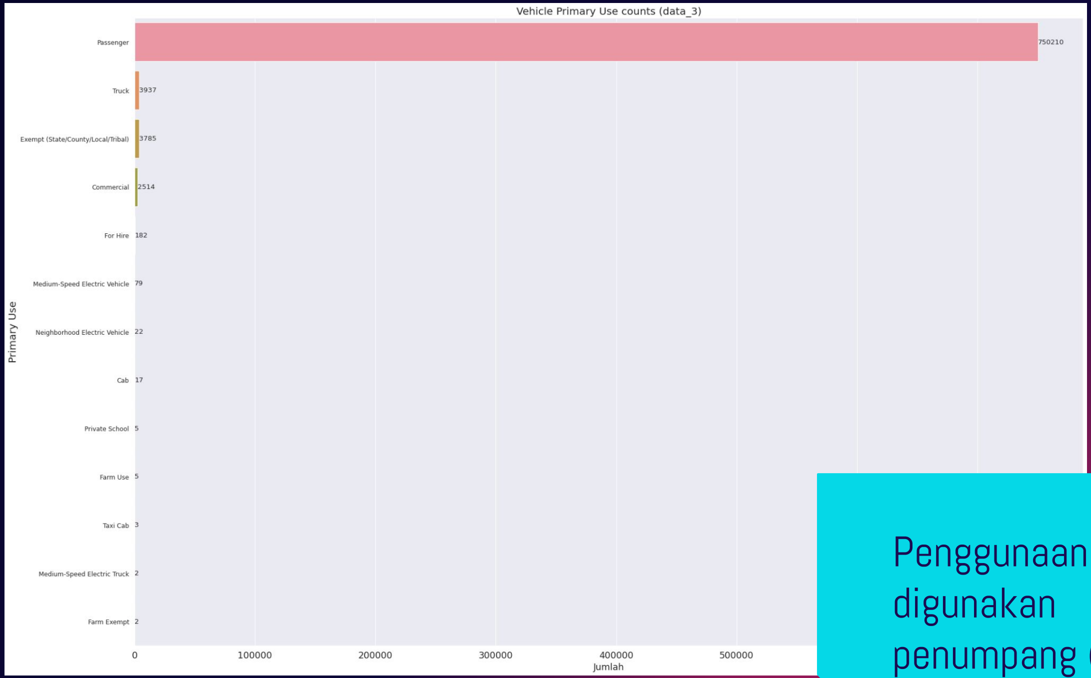
Didapatkan bahwa pada 4 merek terpopuler pertama, Tesla, Nissan, Chevrolet, dan Ford lebih banyak unit yang sudah digunakan dibanding dengan yang baru.

Urutan Model Terpopuler Tesla



Model 3, Model Y, Model S, Model X, Roadster.

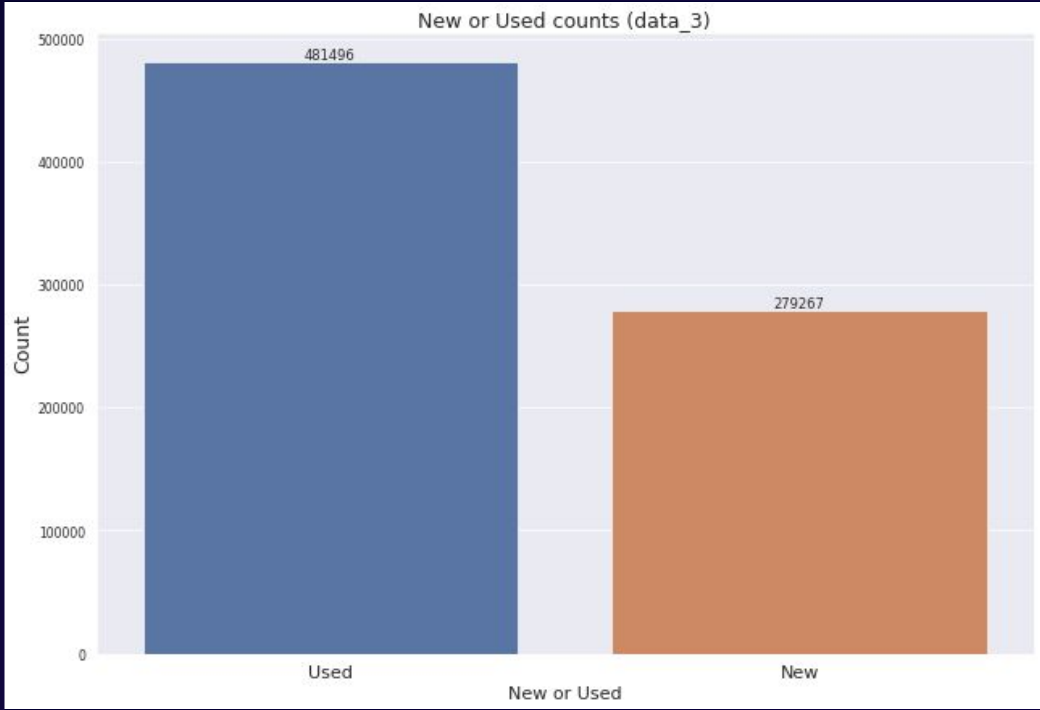
Vehicle Primary Use



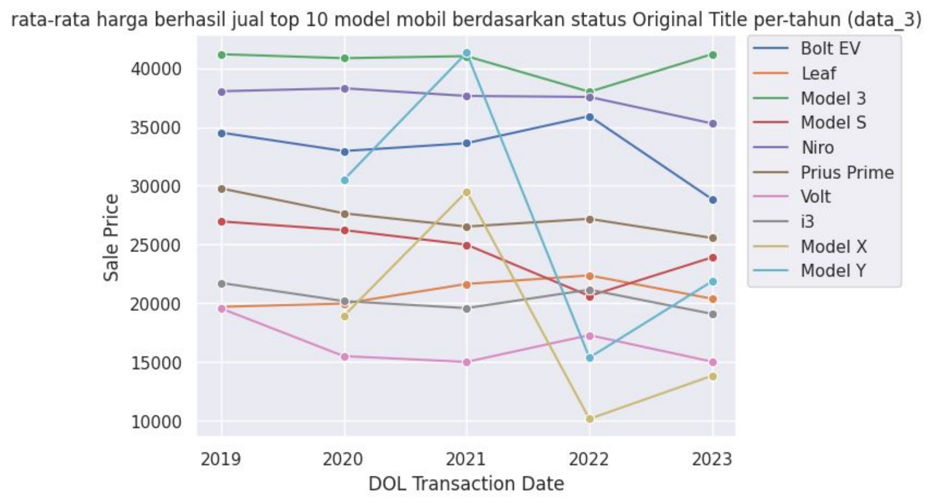
Penggunaan EV di US paling banyak digunakan untuk keperluan menampung penumpang (passengers)

Kondisi Kendaraan EV

Kendaraan EV yang dalam kondisi baru **lebih sedikit** dibandingkan yang sudah pernah terpakai.

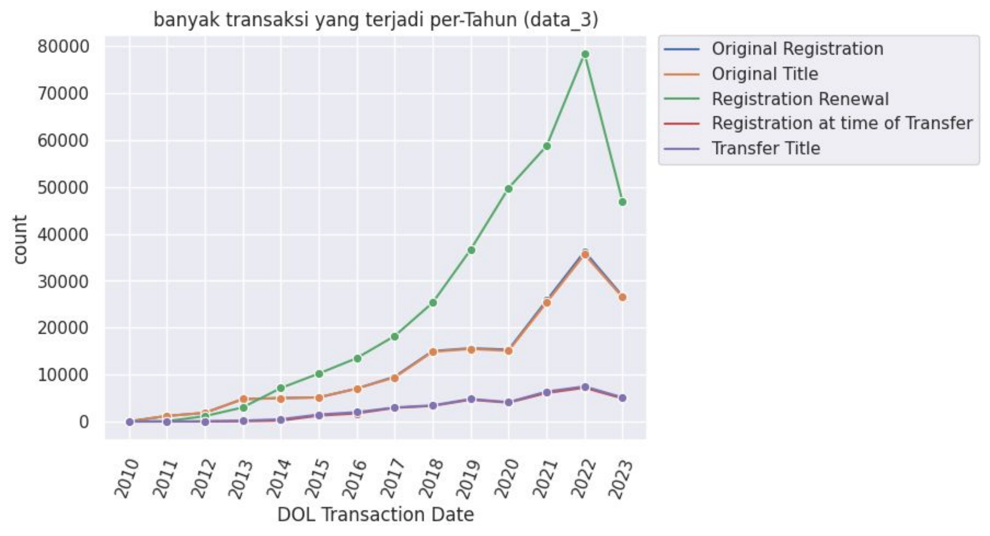


Rata-rata harga jual top 10 model mobil per-tahun



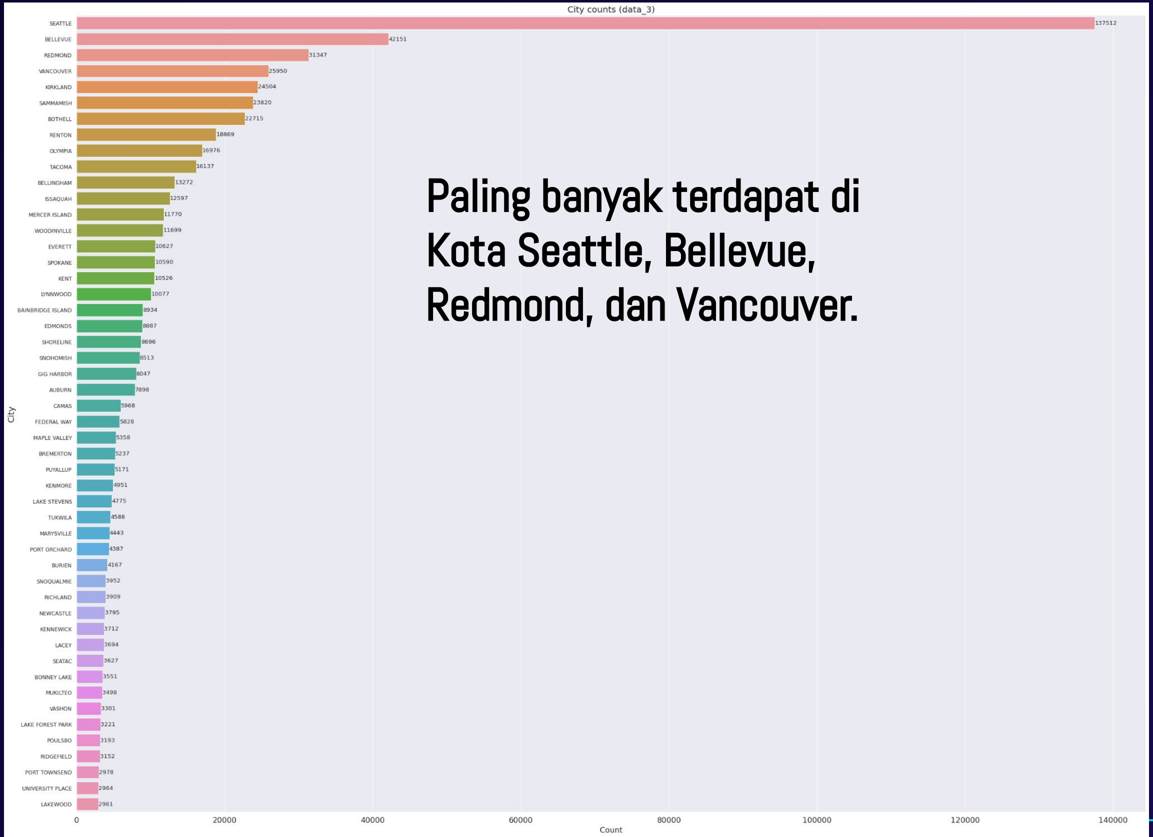
Model 3 memiliki rata-rata harga jual tertinggi, disusul dengan model Niro, dan Bolt EV.

Banyaknya Transaksi yang terjadi Pertahun



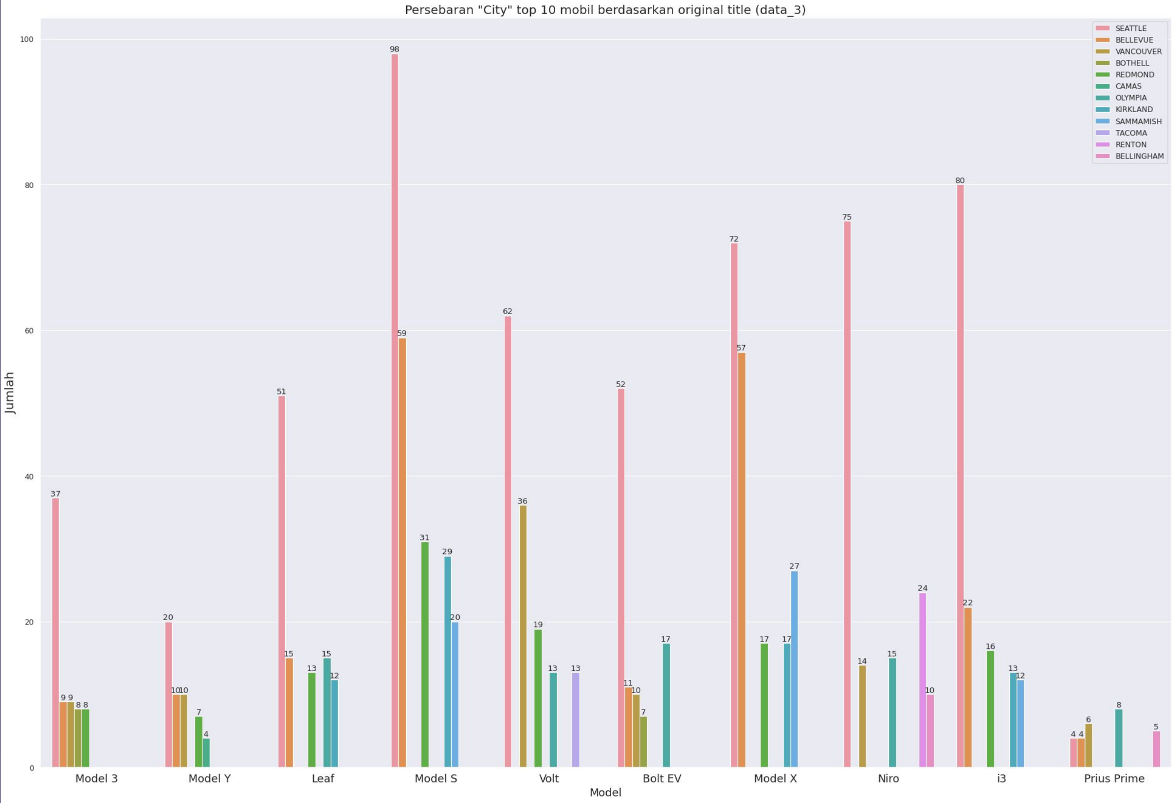
Transaksi terbanyak per tahun adalah Registration Renewal.

Persebaran EV berdasarkan Kota

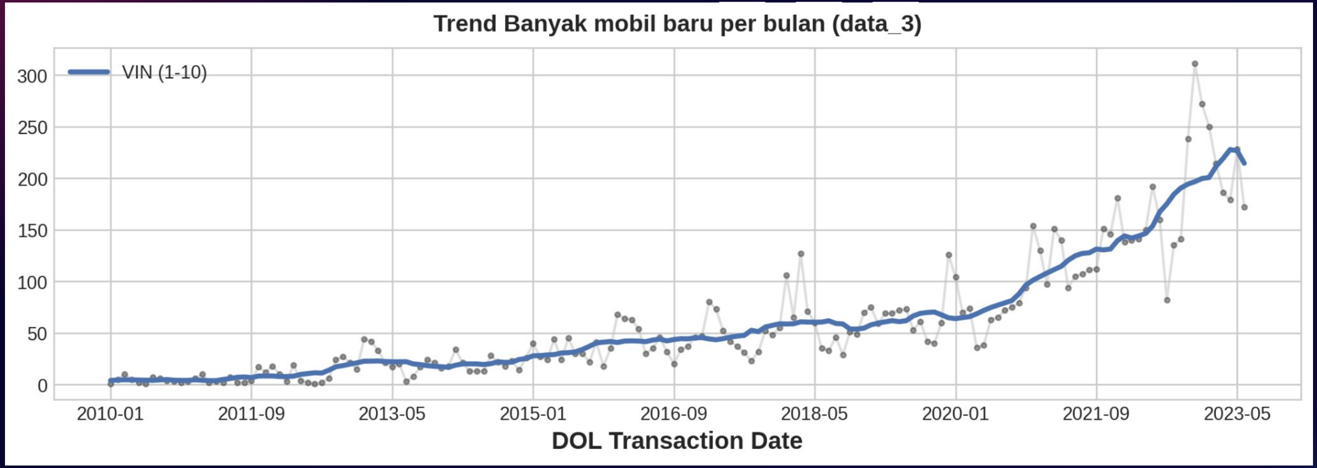


Paling banyak terdapat di Kota Seattle, Bellevue, Redmond, dan Vancouver.

Persebaran Kota Pada Top 10 Model EV Terpopuler



Trend Banyak Mobil Baru per Bulan



Feature Engineering

```
[ ]  tes = data_3.copy()
     tes = tes.drop_duplicates(subset=['VIN (1-10)'])
     tes['DOL Transaction Date'] = tes['DOL Transaction Date'].dt.to_period("M")
     tes = tes.groupby(['DOL Transaction Date'])[['VIN (1-10)']].count()
     tes = tes.to_timestamp()
```

	VIN (1-10)
DOL Transaction Date	
2010-01-01	1
2010-02-01	5
2010-03-01	10
2010-04-01	5
2010-05-01	2
...	...
2023-02-01	214
2023-03-01	186
2023-04-01	179
2023-05-01	228
2023-06-01	172

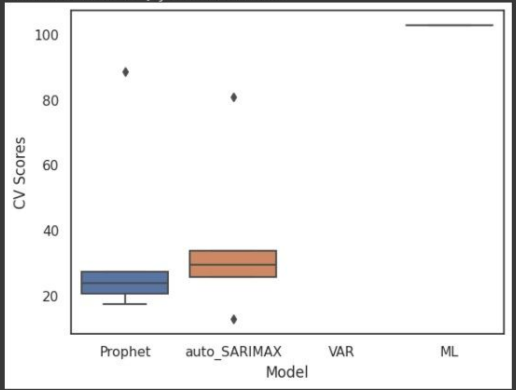
162 rows × 1 columns

Modelling and Evaluation

```python
ts_column = 'DOL Transaction Date'
target = 'VIN (1-10)'

model = auto_timeseries(
            forecast_period=36,
            score_type='rmse',
            time_interval='M',
            non_seasonal_pdq=None, seasonality=False,
            seasonal_period=None,
            model_type='best',
            verbose=2)

model.fit(
            traindata=tes,
            ts_column=ts_column,
            target=target,
        )
```

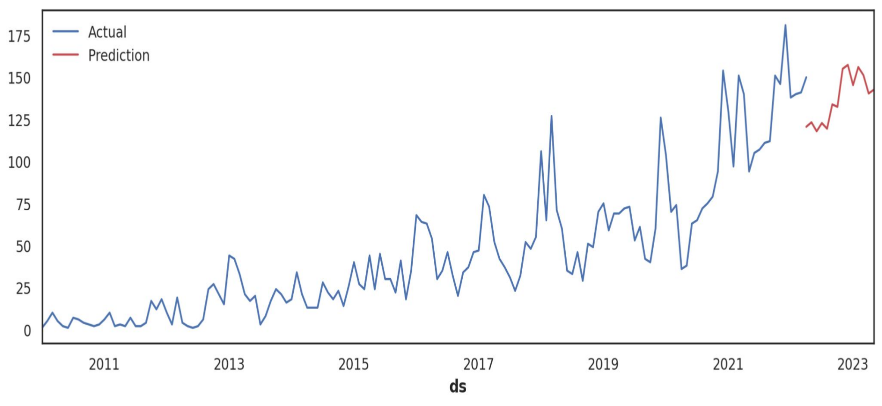
```python
print(model.get_leaderboard())
model.plot_cv_scores()
```

```
          name        rmse
0      Prophet   35.607266
1  auto_SARIMAX   36.538155
3           ML  102.664630
2          VAR         inf
<Axes: xlabel='Model', ylabel='CV Scores'>
```



Prediction

Dapat dilihat bahwa dari model prediksi yang telah dibuat, pertumbuhan EV di masa mendatang akan terus mengalami peningkatan.



Hyperparameter Tuning

```
def mean_absolute_percentage_error(y_true, y_pred):
    y_true, y_pred = np.array(y_true), np.array(y_pred)
    return np.mean(np.abs((y_true - y_pred) / y_true)) * 100
```

```
[180] from sklearn.model_selection import ParameterGrid
      params_grid = {'seasonality_mode':('multiplicative','additive'),
                     'changepoint_prior_scale':[0.1,0.2,0.3,0.4,0.5],
                    'n_changepoints' : [100,150,200]}
      grid = ParameterGrid(params_grid)
      cnt = 0
      for p in grid:
          cnt = cnt+1

      print('Total Possible Models',cnt)
```

```
Total Possible Models 30
```

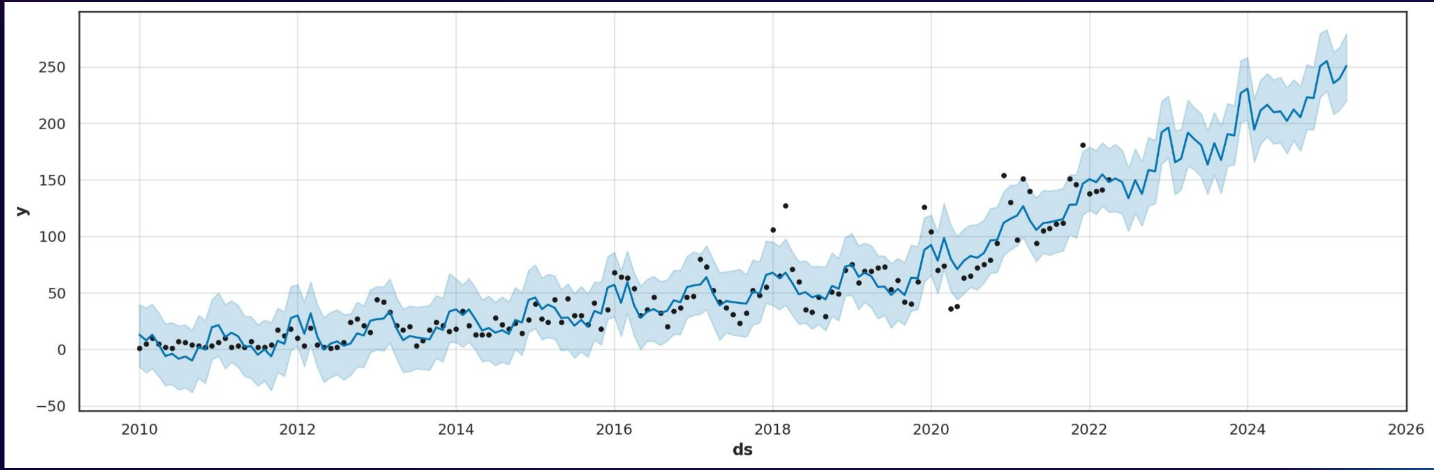
```
[211] strt='2022-04-28'
      end='2023-06-30'

      model_parameters = pd.DataFrame(columns = ['RMSE','Parameters'])
      for p in grid:
          test = pd.DataFrame()
          print(p)
          train_model =Prophet(changepoint_prior_scale = p['changepoint_prior_scale'],
                               n_changepoints = p['n_changepoints'],
                               seasonality_mode = p['seasonality_mode'],
                               weekly_seasonality=False,
                               daily_seasonality = False,
                               yearly_seasonality = True,
                               interval_width=0.95)
          train_model.fit(X_tr)
          train_forecast = train_model.make_future_dataframe(periods=14, freq='M',include_history = False)
          train_forecast = train_model.predict(train_forecast)
          test=train_forecast[['ds','yhat']]
          Actual = df[(df['ds']>strt) & (df['ds']<=end)]
          MAPE = mean_absolute_percentage_error(Actual['y'],abs(test['yhat']))
          print('Mean Absolute Percentage Error(MAPE)------------------------------------',MAPE)
          model_parameters = model_parameters.append({'MAPE':MAPE,'Parameters':p},ignore_index=True)
```

```
parameters = model_parameters.sort_values(by=['MAPE'])
parameters = parameters.reset_index(drop=True)
parameters.head()
```

	RMSE	Parameters	MAPE
0	NaN	{'changepoint_prior_scale': 0.5, 'n_changepoin...	20.580553
1	NaN	{'changepoint_prior_scale': 0.5, 'n_changepoin...	20.717934
2	NaN	{'changepoint_prior_scale': 0.5, 'n_changepoin...	20.717934
3	NaN	{'changepoint_prior_scale': 0.4, 'n_changepoin...	21.024251
4	NaN	{'changepoint_prior_scale': 0.4, 'n_changepoin...	21.024251

Hasil Hyperparameter Tuning

```
[214] print("Best Parameter",parameters['Parameters'][0])
      print("MAPE Score: ",parameters['MAPE'][0])

      Best Parameter {'changepoint_prior_scale': 0.5, 'n_changepoints': 100, 'seasonality_mode': 'multiplicative'}
      MAPE Score:  20.58055283392801
```



Conclusion & Recommendation

Terlihat dari hasil prediksi time series kami bahwa pembelian kendaraan EV terus meningkat tiap tahunnya.

Terlihat bahwa pembelian meningkat di akhir tahun 2019 dimana terjadi pemberlakuan pembebasan pajak bagi mobil EV yang memenuhi kriteria.

Kota dengan EV terbanyak adalah seattle dengan 1214 EV Charging Stations dan kota dengan EV Tersedikit adalah Lakewood dengan 20 EV Charging Stations

Tipe kendaraan elektronik yang paling banyak terjual adalah tipe BEV (battery electric vehicle)

Penggunaan EV terbanyak digunakan untuk membawa penumpang

Rekomendasi kami untuk para produsen mobil EV adalah terus mendorong pemerintah untuk memperbanyak insentif terhadap mobil EV, memperbanyak EV Charging Stations pada kota-kota agar dapat menjangkau pasar yang belum terjangkau sebelumnya, memproduksi lebih banyak mobil bertipe BEV (battery electric vehicle) dan tetap mempertahankan produksi mobil dengan tujuan membawa penumpang

Thank You

Calculus Clan