

DATA SCIENCE ACADEMY

A Report from
Team Sukinep

Disusun Oleh:



Joshua Galilea



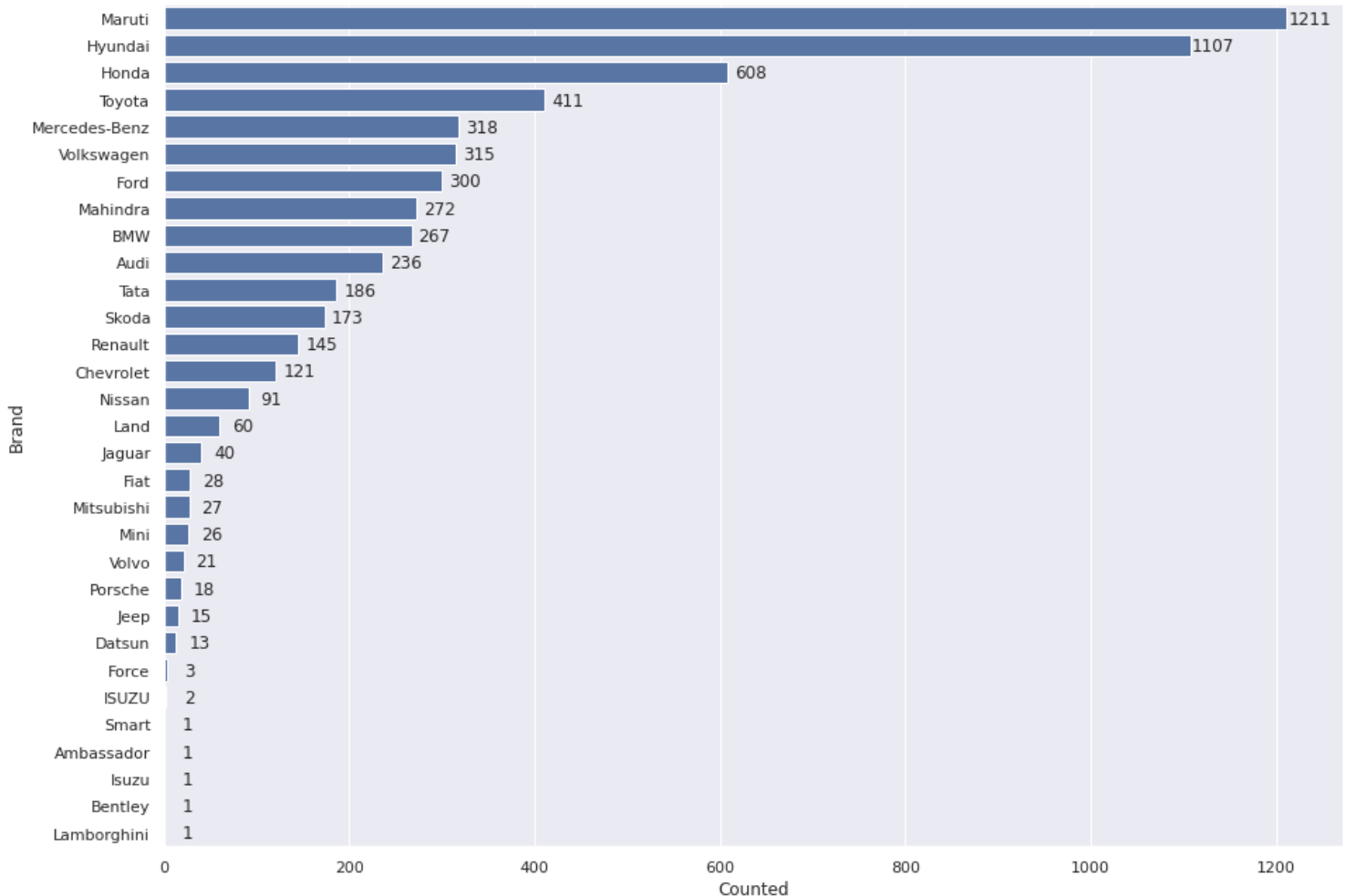
M. Ahsan Athallah

LATAR BELAKANG

Laporan ini ditulis dengan tujuan untuk memenuhi persyaratan pendaftaran **Data Science Academy** yang diadakan oleh **COMPFEST 12**. Data yang digunakan dalam analisis ini adalah data yang dibagikan oleh panitia COMPFEST 12 dalam tautan [ini](#). Data tersebut berisikan 6019 baris data spesifikasi mobil bekas yang ada di India. Dalam laporan ini terdapat jawaban dari 10 soal yang diberikan oleh tim COMPFEST 12. Tidak hanya itu, pada laporan ini juga dituliskan beberapa *insight*, mengenai mobil bekas, yang turut kami sisipkan.

JAWABAN SOAL

[Nomor 1] Merek mobil apa saja yang tersedia dan ada berapa banyak mobil untuk tiap merek tersebut?



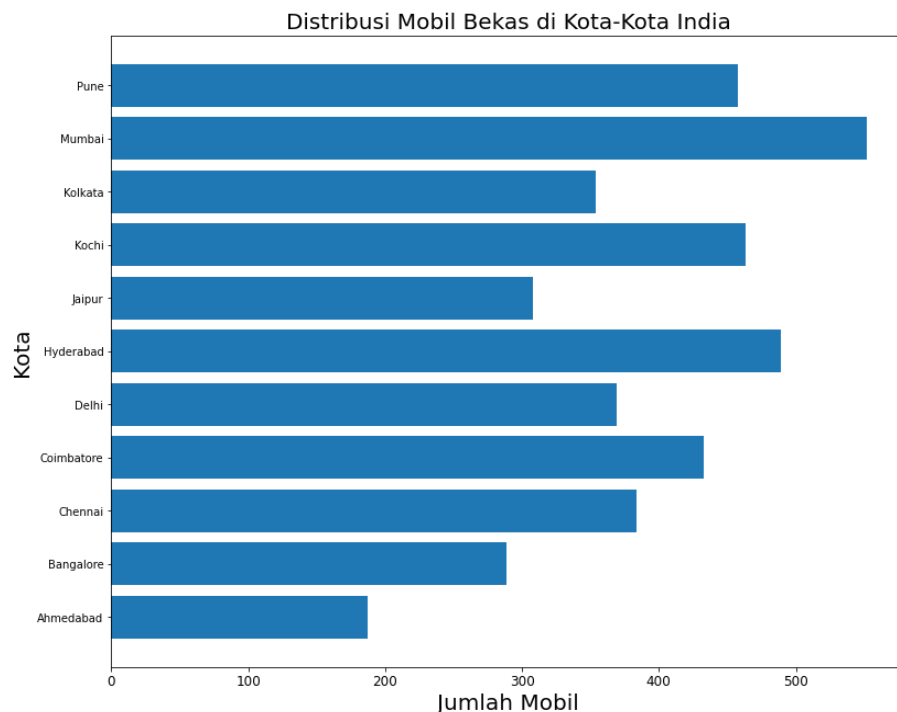
Dari visualisasi data diatas kami menemukan **31 merek mobil bekas** di India. **Tiga merek mobil bekas** yang **paling banyak beredar** adalah **Maruti** (1211 unit), **Hyundai** (1107 unit) dan **Honda** (608 unit). Untuk memastikan kebenaran data, kami turut menggunakan data penjualan mobil bekas yang dirilis oleh **Cars 24** (situs jual beli mobil bekas di India) dan menemukan **penjualan mobil bekas tertinggi** pada merek **Maruti** (35.51%), **Hyundai** (23.64%), dan **Honda** (8.64%)^[1]. Urutan ini tentu sesuai dengan urutan jumlah mobil bekas (per merek) dari data yang telah kami olah.

[Nomor 2] Kota apa yang memiliki mobil bekas paling banyak?

Berdasarkan data, **Kota Mumbai** adalah kota yang paling banyak memiliki mobil bekas, yakni **sebanyak 552 mobil bekas**. Angka tersebut diperoleh dengan cara, **mengelompokkan data mobil bekas berdasarkan kota** yang tersedia.

```
train = train.groupby('Location')['Name'].nunique()
print(train)
```

Location	Hyderabad	489
Ahmedabad	Jaipur	308
Bangalore	Kochi	463
Chennai	Kolkata	354
Coimbatore	Mumbai	552
Delhi	Pune	458



Hal ini juga dibenarkan oleh situs **The Times of India**, bahwa **Kota Mumbai** adalah kota dengan **penjualan mobil bekas terbanyak** atau dapat diartikan kalau kota ini memiliki populasi mobil bekas terbanyak^[2].

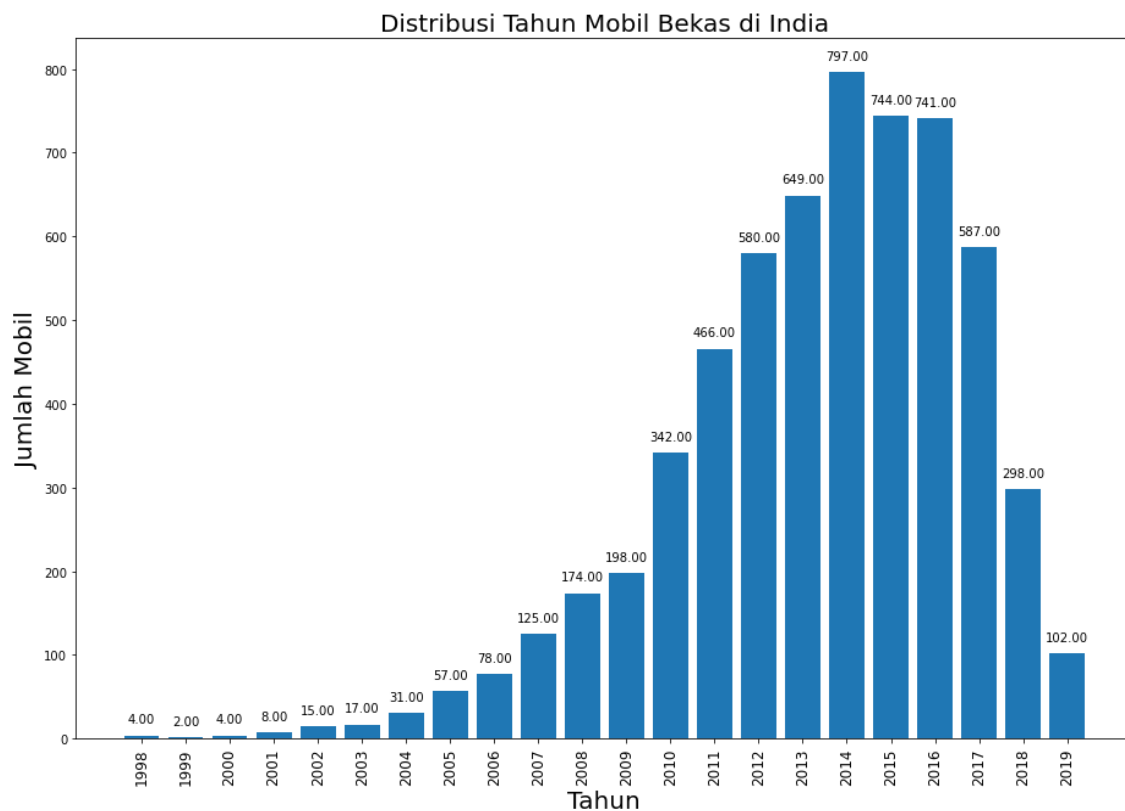
[Nomor 3] Bagaimana distribusi tahun edisi mobil-mobil bekas tersebut?

Untuk mengetahui persebaran mobil bekas berdasarkan tahun edisi nya dapat dilakukan **pengelompokan data mobil bekas berdasarkan kolom 'Year'**.

```
# Pengelompokan data berdasarkan kolom 'year'  
train.groupby('Year').count()
```

Lantas, dilakukan visualisasi data dan ditemukan hasil sebagai berikut:

- Mobil bekas yang **paling banyak beredar** adalah mobil bekas tahun **2014**, atau sebanyak **797 mobil**.
- Mobil bekas yang **paling sedikit beredar** adalah mobil bekas tahun **1999**, atau sebanyak **2 mobil**.



[Nomor 4] Berapa banyak mobil yang memiliki total jarak pemakaian di bawah 100.000 kilometer?

Langkah pertama yang kami ambil adalah melakukan penyortiran (*filter*) data mobil bekas yang memiliki total jarak pemakaian dibawah 100.000 kilometer. Berikut adalah cara kami menerapkan *filter* yang dimaksud.

```
[105] df.Name[df.Kilometers_Driven < 100000].count()
5470
```

Dari proses penyortiran tersebut, **kami temukan bahwa terdapat 5470 mobil bekas** dengan total jarak pemakaian kurang dari 100.000 kilometer.

[Nomor 5] Pada batas berapa kilometer total jarak pemakaian bisa dikategorikan sebagai rendah atau tinggi? Sertakan argumen yang mendukung jawaban.

Agar kami bisa memilah antara mobil bekas dengan jarak pemakaian yang rendah dan jarak pemakaian yang tinggi, kami memutuskan untuk menetapkan aturan pengkategorisasian jarak pemakaian, yaitu sebagai berikut.

Kilometer	Range
Rendah	≤ 34000
Sedang	$34000 < x \leq 73000$
Tinggi	> 73000

Aturan yang kami buat berlandaskan data **kuartil 1 (Q1)** dan data **kuartil 3 (Q3)** yang kami temukan dari keseluruhan kolom '**Kilometers_Driven**' sebagaimana berikut.

```
des = train.Kilometers_Driven.describe()
Q1,Q3 = des['25%'],des['75%']
Q1,Q3
```

```
(34000.0, 73000.0)
```

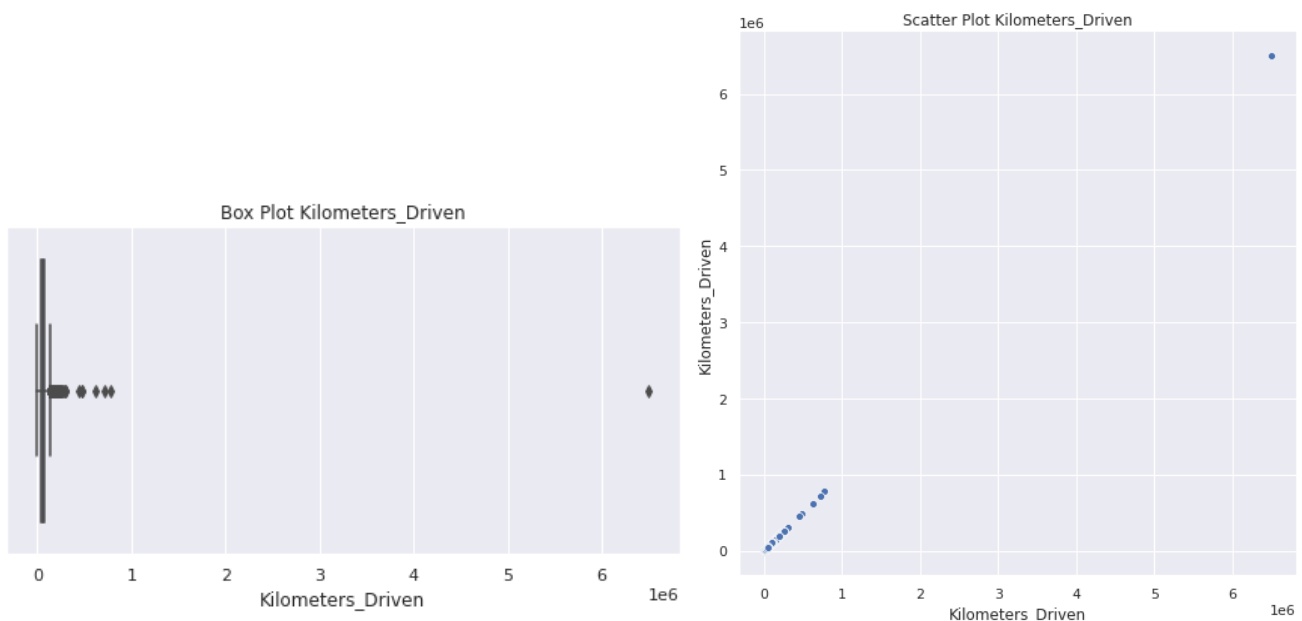
Kemudian, kami menerapkan aturan tersebut untuk melakukan kategorisasi terhadap jarak pemakaian. Lantas, kami menemukan sebanyak **1492 mobil bekas dengan kategori jarak pemakaian tinggi** dan **1524 mobil bekas dengan jarak pemakaian rendah**.

```
df.groupby('KD_category').count().Name
```

```
0 KD_category
  Rendah    1524
  Sedang    3003
  Tinggi    1492
  Name: Name, dtype: int64
```

[Nomor 6] Apakah terdapat outlier pada kolom Kilometers_Driven?

Jawabannya adalah **Ya**, terdapat **satu data pencilan** yang sangat berbeda dari data yang lain. Berdasarkan informasi yang didapat dari boxplot dan scatter plot.



Terlihat bahwa ada data yang **sangat jauh berbeda** dengan data yang lain. Pada kasus ini data pencilan itu jauh lebih besar daripada data lain. Dengan memanfaatkan Z-Score atau Standar Deviasi dari persebaran data pada kolom 'Kilometers_Driven', dapat ditemukan sebuah data yang berada sangat jauh dari data lainnya. Data tersebut berada pada index **2328** dengan nilai **6.500.000**.

```
z = np.abs(stats.zscore(train.Kilometers_Driven))
threshold = 7
outlier = np.where(z > threshold)
for o in outlier:
    print(train.Kilometers_Driven[o])
```

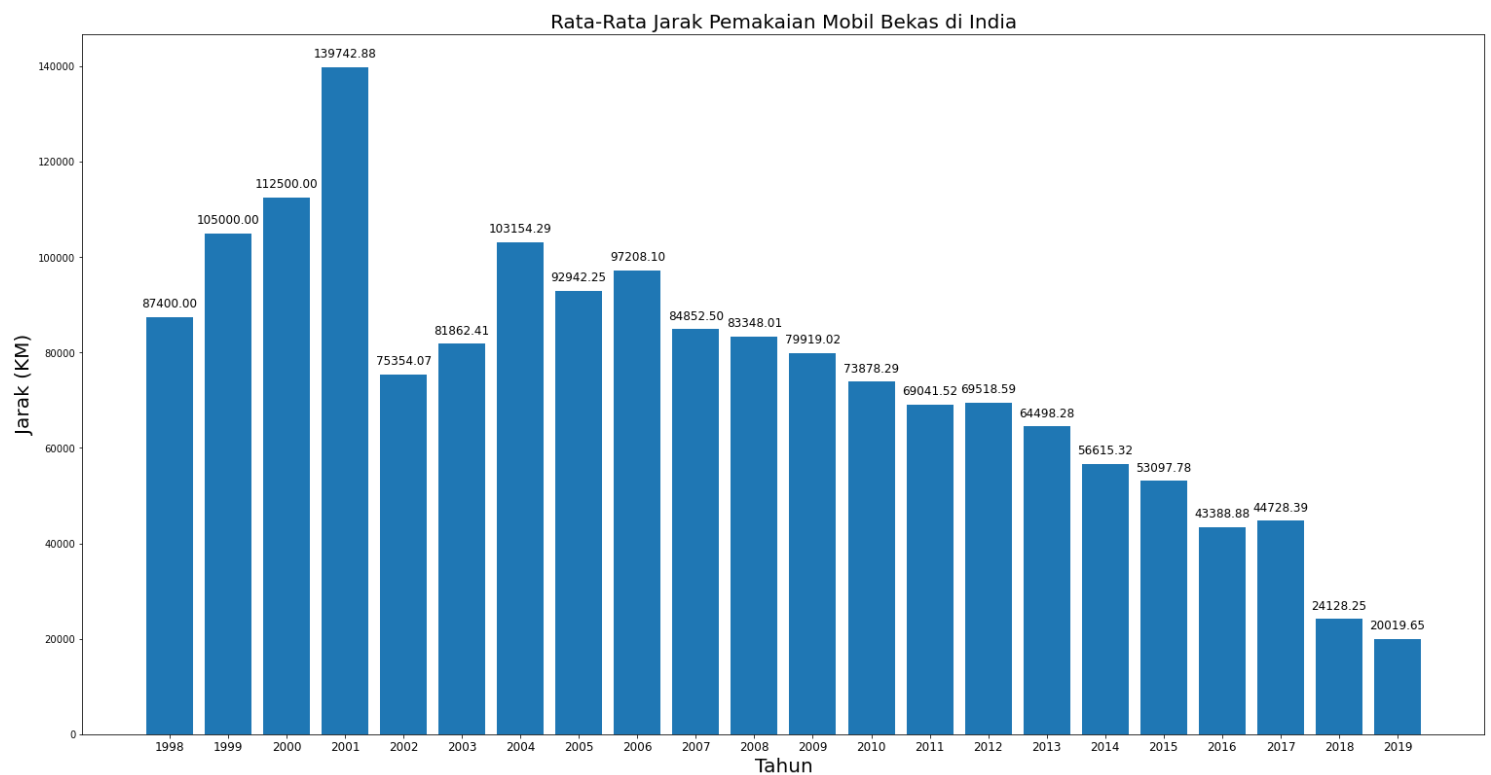
```
340      775000
1860     720000
2328    6500000
Name: Kilometers_Driven, dtype: int64
```

**[Nomor 7] Apakah tahun pembuatan mobil berpengaruh terhadap total jarak pemakaian?
Sertakan argumen yang mendukung jawaban.**

Kami mulai dengan cara mengelompokkan data berdasarkan kolom **'Kilometers_Driven'** masing-masing mobil berdasarkan tahun produksinya, seperti berikut.

```
#Pengelompokan data kolom 'Kilometer_Driven'  
#berdasarkan kolom 'Year'  
train.groupby('Year').mean().Kilometers_Driven
```

Lantas, dilakukan visualisasi data dan kami menemukan **jarak pemakaian mobil bekas yang cenderung menurun** dibanding tahun sebelumnya, dimulai dari mobil bekas yang diproduksi tahun 2007. Walaupun tidak seluruh data menunjukkan penurunan jarak tempuh, tetap dapat disimpulkan bahwa, tahun mobil bekas memengaruhi jarak pemakaiannya. Semakin **'muda'** usia mobil bekas, maka mobil tersebut **berkemungkinan** memiliki jarak tempuh yang rendah.

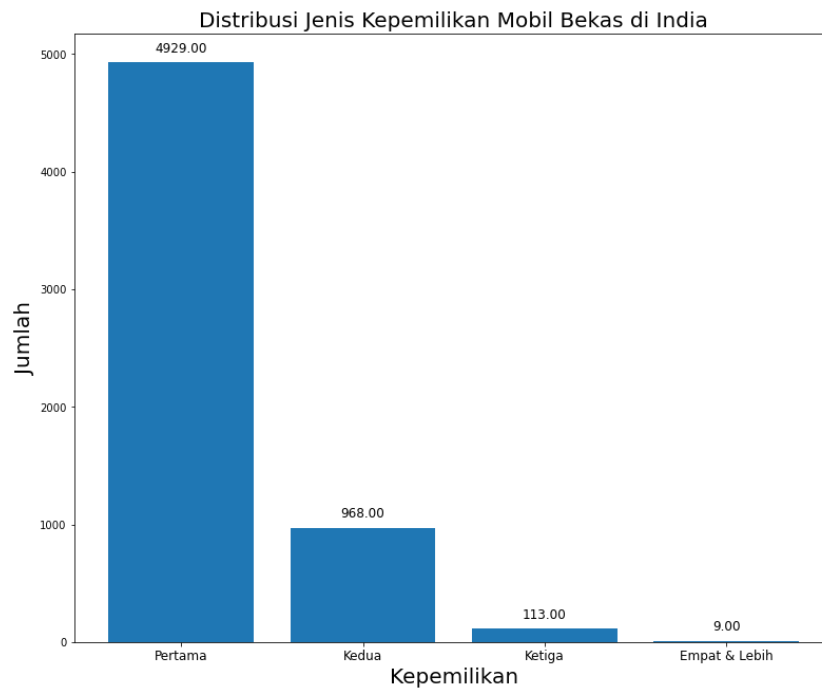


[Nomor 8] Berapa banyak mobil yang merupakan kepemilikan ketiga atau lebih?

Untuk mengetahui banyaknya mobil bekas kepemilikan ketiga atau lebih, pertama-tama dilakukan pengelompokan data berdasarkan kolom **'Owner_Type'**.


```
# Pengelompokan data berdasarkan 'Owner_Type'
train.groupby('Owner_Type').count().Name
```

Lantas, dilakukan visualisasi dan ditemukan jawaban untuk mobil bekas yang merupakan kepemilikan ketiga atau lebih yaitu **sebanyak 121 mobil**.



[Nomor 9] Tipe bahan bakar apa yang memiliki mileage (konsumsi bahan bakar) paling hemat?

Untuk mengetahui tipe (jenis) bahan bakar terhemat sekaligus yang terboros dapat ditemukan dengan mengikuti tahapan berikut:

1. Pengelompokan data mobil sesuai dengan jenis bahan bakarnya

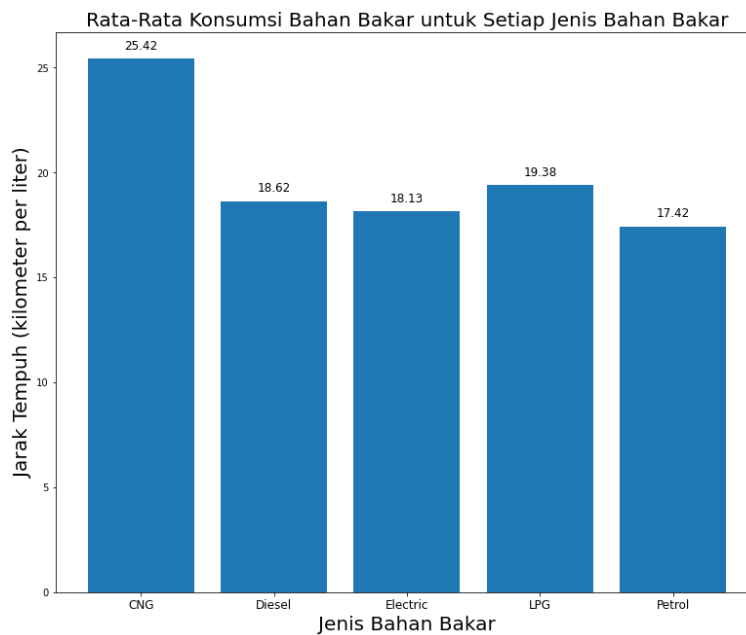
```
# Pengelompokan data berdasarkan jenis bahan bakar
for i in range(len(use)): #use merupakan array dua dimensi berisi 'Fuel_Type' dan 'Mileage'
    if use[i][0] == 'CNG':
        cng.append(use[i])
    elif use[i][0] == 'Diesel':
        diesel.append(use[i])
    elif use[i][0] == 'Electric':
        electric.append(use[i])
    elif use[i][0] == 'LPG':
        lpg.append(use[i])
    elif use[i][0] == 'Petrol':
        petrol.append(use[i])
```

- Setelah dilakukan pengelompokan, lantas dilakukan **penghitungan nilai rata-rata** jarak tempuh untuk masing-masing jenis bahan bakar. Namun, diperlukan aksi pengubahan tipe data string ke float (pecahan) untuk data pada kolom 'Mileage'.

```
#Mengubah tipe data String menjadi Float
def convertToFloat(cng):
    mil = []
    for i in range(len(cng)):
        if cng[i][1] != 0.0:
            temp = cng[i][1].split()
            temp2 = float(temp[0])
            mil.append(temp2)
    return findAverageMileage(mil)

#Mencari nilai average jenis bahan bakar bersangkutan
def findAverageMileage(mil):
    avg = 0
    for i in range(len(mil)):
        avg += mil[i]
    return avg / len(mil)
```

3. Setelah ditemukan rata-rata, barulah dapat ditentukan jenis bahan bakar terhemat dan jenis bahan bakar terboros. Berikut adalah grafik yang menunjukkan rata-rata konsumsi bahan bakar untuk setiap jenis bahan bakar.



Dari data pada grafik diatas, terjawab bahwa jenis bahan bakar yang paling hemat adalah **CNG**.

[Nomor 10] Apa saja faktor-faktor yang mempengaruhi harga mobil bekas di India?

Sertakan argumen yang mendukung jawaban.

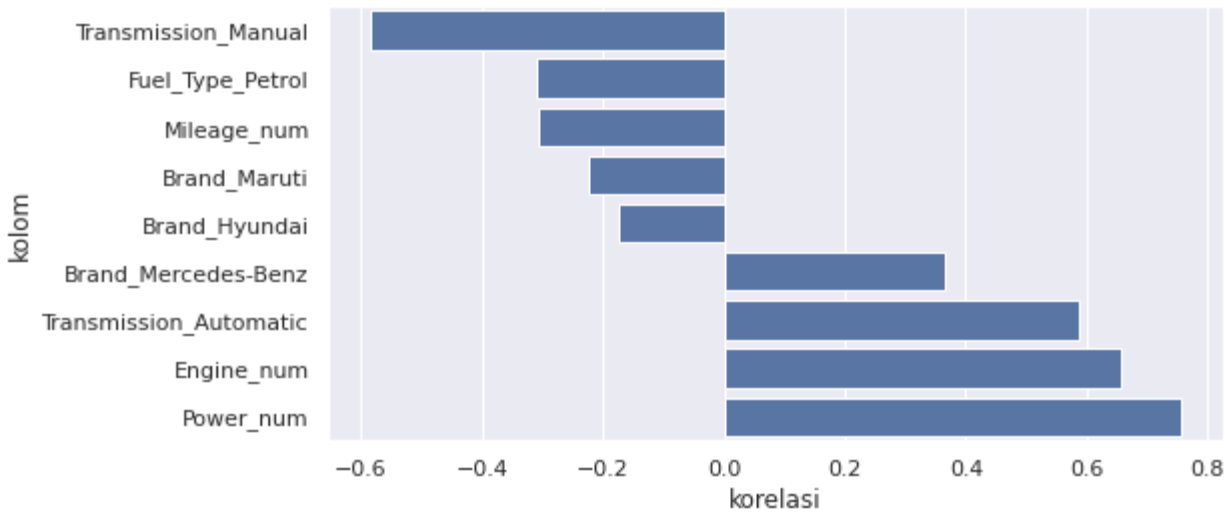
Jawaban untuk pertanyaan ini akan menggunakan **korelasi Pearson** untuk menentukan apakah suatu informasi berpengaruh terhadap target. Korelasi Pearson adalah sebuah metodologi untuk mengukur kekuatan asosiasi linear antara dua variabel^[3]. Korelasi positif menandakan bahwa peningkatan maupun penurunan nilai dari setiap kolom akan sejalan dengan kolom target. Begitu juga sebaliknya untuk korelasi negatif. Dalam kasus ini, kami akan membandingkan antara beberapa variabel dengan variabel target yakni kolom **price**. Beberapa hal yang kami persiapkan sebelum menghitung korelasi adalah:

1. Menambahkan kolom "Brand" dengan cara yang sama dengan soal nomor satu.
2. Mengisi data yang nan atau null dengan rata-rata dari setiap kolom yang berisi data kontinu.

3. Memisahkan nilai dan satuan dari kolom-kolom yang memiliki nilai kontinu tetapi bentuk datanya adalah *string*. Kolom yang berisikan nilai dari data kontinu akan bernama “(Nama Kolom Asli)_num”
4. Memecah semua kolom yang berisikan data kategori menjadi sebuah kolom sendiri (*One Hot Encoding*) untuk mendapatkan detail pengaruh setiap kategori terhadap harga dari mobil bekas. Nama kolom baru akan memiliki format berikut “(Nama Kolom Asli)_(Kategori)”. Kolom hasil One Hot Encoding ini nantinya akan berisi angka 0 dan

Jawaban:

Berdasarkan hasil dari *Pearson Correlation* yang telah dilakukan pada 6019 Baris dan 81 Kolom dapat lihat bahwa, terdapat 9 variabel (faktor) yang mempengaruhi harga mobil bekas India. Faktor yang paling mempengaruhi harga mobil bekas adalah variable ‘**Transmission_manual**’ (transmisi manual dapat mengurangi harga mobil dengan signifikan) dan variable ‘**Power_Num**’ (semakin tinggi horsepower, semakin mahal harga mobil), sebagaimana ditampilkan pada grafik dibawah ini.



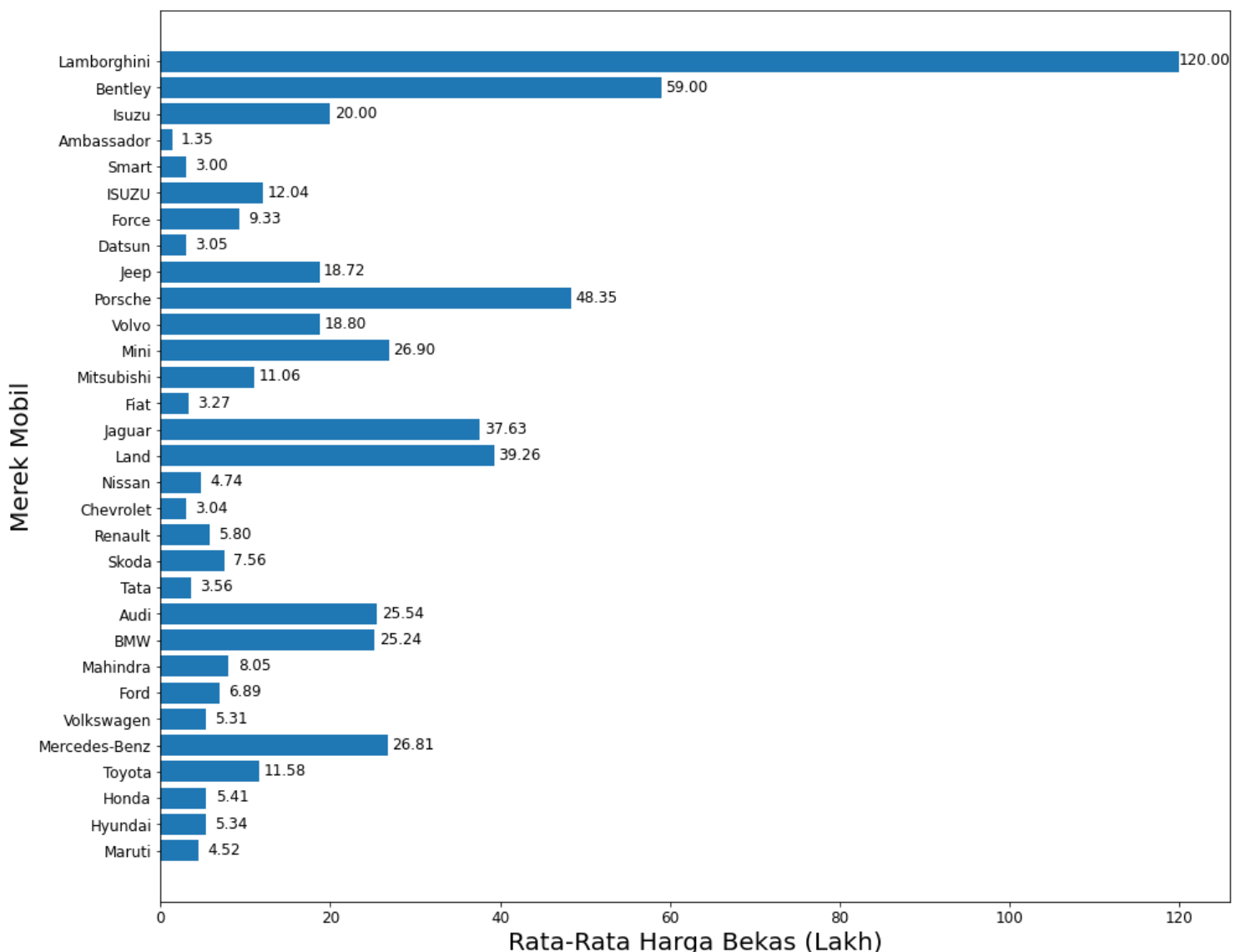
HASIL ANALISIS TAMBAHAN

Mengapa merek mobil Maruti dan Hyundai beredar paling banyak di kalangan masyarakat India?

Dari data yang kami olah, ditemukan merek Maruti dan Hyundai merupakan dua merek mobil bekas yang **paling diminati oleh masyarakat india**. Hal ini disebabkan oleh **jumlah mobil bekas** kedua merek tersebut **beredar paling banyak** dibandingkan merek mobil bekas lainnya.



Pertanyaan kami adalah **apa yang menyebabkan mereka digemari masyarakat India?** Pertama-tama, kami mencari **harga rata-rata** (tanpa pembagian tahun mobil bekas) kedua merek mobil tersebut dan menemukan data seperti dibawah ini.



Ternyata merek Maruti (4.52 lakh) dan Hyundai (5,34 lakh) **bukanlah** merek mobil bekas dengan **rata-rata harga terendah**. Melainkan terdapat merek Ambassador (1.35 lakh), Smart (3.00 lakh), dan Chevrolet (3.04 lakh), yang merupakan merek dengan harga mobil bekas terendah. Hal ini membuat kami semakin penasaran, lantas kami terpikir untuk **membagi rata-rata harga setiap merek mobil untuk tiga tahun terakhir** yakni, tahun 2017, 2018, dan 2019. Lalu, inilah yang kami temukan.

- **Mobil Bekas merek Ambassador**, tercatat **hanya memiliki 1 unit mobil bekas** dan itu pun diproduksi **tahun 2003**. Bukan hanya lawas, kemungkinan alasan yang menyebabkan mobil ini tidak diminati masyarakat India adalah, mobil bekas merek Hyundai dan Maruti menawarkan **lebih banyak variasi** dan **kondisi** (tahun, jarak tempuh, dan harga). Hal ini nampaknya juga berlaku untuk merek mobil **Smart**.

Ambassador	1	Ambassador 2003 1.35	Smart	1	Smart 2008 3.0
------------	---	----------------------	-------	---	----------------

- Ditemukan harga mobil bekas Maruti dan Hyundai, tiga tahun terakhir, **tidak sepenuhnya lebih rendah** dari harga mobil bekas lainnya. Walaupun demikian, kedua merek populer tersebut memiliki **tahun yang lebih beragam (lengkap)** ketimbang merek lainnya.

Harga merek mobil Maruti dan Hyundai

```
Maruti 2017 6.749015151515152
Maruti 2018 7.486233766233766
Maruti 2019 6.406842105263158
=====
Hyundai 2017 7.800569105691059
Hyundai 2018 8.384218749999997
Hyundai 2019 9.818947368421052
=====
```

Harga merek mobil lain yang dibandingkan

Datsun 2017 3.453333333333333	Volkswagen 2017 8.061818181818182
Datsun 2018 3.54	Volkswagen 2018 7.968999999999999
Datsun 2019 0.0	Volkswagen 2019 5.68
=====	=====
Nissan 2017 5.85	Tata 2017 5.5200000000000005
Nissan 2018 4.0	Tata 2018 7.0
Nissan 2019 0.0	Tata 2019 9.014
=====	=====
Chevrolet 2017 7.75	Mahindra 2017 9.783333333333331
Chevrolet 2018 9.87	Mahindra 2018 12.158461538461536
Chevrolet 2019 0.0	Mahindra 2019 12.366666666666665
=====	=====

Artinya **tidak benar** kalau kedua merek mobil tersebut memiliki harga rata-rata yang lebih rendah diantara merek mobil lainnya.

Tidak sampai disitu, kami pun tetap mencoba untuk mencari faktor lain yang menyebabkan kedua merek mobil tersebut paling banyak beredar di kalangan masyarakat India. **Faktor kedua** yang kami coba telusuri adalah **apakah mobil bekas merek Maruti dan Hyundai memiliki rata-rata *mileage* yang rendah?** Lantas inilah yang kami temukan.

```

Brand - Average Mileage
Maruti - 21.505681255161033
Hyundai - 18.729105691056937
Datsun - 21.590769230769226
Nissan - 19.361758241758256
Chevrolet - 18.40115702479339
Volkswagen - 18.258888888888888
Tata - 19.83973118279571

```

Tetap, mobil merek Maruti dan Hyundai **bukanlah** mobil yang paling ‘irit’ bahan bakar. Jadi, dapat dikatakan bahwa kedua merek tersebut **tidak memiliki rata-rata *mileage* yang rendah**.

Lantas apa yang menjadi, kelebihan dua mobil ini? Kami memutuskan untuk melakukan **pencarian keluar dari data yang kami olah**, karena dan mencoba mengunjungi beberapa situs otomotif India. Pencarian kami terhenti pada laman **gomechanic.in**. Kami menemukan bahwasanya, kedua Maruti dan Hyundai memiliki ‘emas’-nya masing-masing, berikut adalah kelebihan Maruti dan Hyundai menurut laman tersebut.

Kelebihan Maruti ^[4]	Kelebihan Hyundai ^[5]
<ul style="list-style-type: none"> • Masyarakat India sangat percaya pada merek Maruti yang telah beredar selama lebih dari 40 tahun. • Suku cadang Maruti tersedia cukup luas. Bengkel Maruti dapat ditemukan setiap selisih 5 s.d 6 kilometer di India. 	<ul style="list-style-type: none"> • Mobil Hyundai disematkan banyak fitur, seperti <i>wireless charger, the smart boot, emergency stop signal, ventilated front seats, dan twin-tip mufflers</i>.

Kesimpulan yang dapat diambil dari pertanyaan “**Mengapa merek mobil Maruti dan Hyundai beredar paling banyak di kalangan masyarakat India?**” adalah pada kenyataan kami **tidak menemukan faktor** pada data yang menyebabkan kedua merek ini memiliki jumlah peredaran paling tinggi di India, jika ternyata dapat ditemukan faktor yang dimaksud kami memohon maaf sebab tak terpikir untuk melakukan eksplorasi data yang lebih. Namun, walaupun kedua merek tersebut tidak memiliki **rata-rata harga termurah** dan tidak memiliki **rata-rata *mileage* (jarak tempuh) yang cukup jauh**, tetap ada faktor eksternal (diluar data), yang lebih manusiawi untuk dipertimbangkan oleh pembeli, seperti fitur, kepercayaan merek, dan layanan purna jual.

DAFTAR PUSTAKA

- [1] Cars24.com. (2017, 1 Maret). *India's Best Selling Second Hand Cars In 2016*. diakses pada 17 Juli 2020, dari <https://www.cars24.com/blog/indias-best-second-hand-cars/>
- [2] Timesofindia.indiatimes.com. (2020, 18 Mei). *Rising share of pre-owned car segment in India*, diakses pada 17 Juli 2020, dari <https://timesofindia.indiatimes.com/blogs/voices/rising-share-of-pre-owned-car-segment-in-india/>
- [3] Sedgwick, P. (2012). *Pearson's Correlation Coefficient*. *Bmj*, 345, e4483.
- [4] Gomechanic.in. (2020, 4 Juni). *10 Reasons Why India Loves Maruti Suzuki Cars*. diakses pada 18 Juli 2020, dari <https://gomechanic.in/blog/indians-love-maruti-suzuki-cars/>
- [5] Gomechanic.in. (2020, 7 Juni). *10 Reasons Why India is in Love with Hyundai Cars*. diakses pada 18 Juli 2020, dari <https://gomechanic.in/blog/why-india-loves-hyundai-cars/>