

Distribusi Restoran Pada Daerah Tertentu

Konten

- [Pendahuluan](#)
- [Tahap 1. Pendahuluan](#)
- [Tahap 2. Import Library](#)
- [Tahap 3. Pra Pemrosesan Data](#)
- [Tahap 4. Analisa Data](#)
 - [4.1 Proporsi Dari Berbagai Jenis Usaha]
 - [4.2 Proporsi Dari Usaha Berbasis Waralaba dan Nonwaralaba]
 - [4.3 Jenis Usaha Yang Berbasis Waralaba]
 - [4.4 Yang Merupakan Ciri-Ciri Dari Jenis Usaha Berbasis Waralaba]
 - [4.5 Rata-Rata Kursi Untuk Setiap Jenis Usaha]
 - [4.6 Memisahkan Nama Alamat]
 - [4.7 Grafik Yang Menampilkan Sepuluh Nama Jalan Teratas]
 - [4.8 Jumlah Jalan Yang Memiliki Satu Restoran]
 - [4.9 Distribusi Jumlah Kursi]
- [Kesimpulan](#)

Pendahuluan

Anda telah memutuskan untuk membuka sebuah kafe kecil di Los Angeles. Kafe Anda bisa dibilang cukup unik, karena Anda "memperkerjakan" robot sebagai pelayan atau pramusaji. Proyek ini memang menjanjikan, namun membutuhkan modal yang besar. Oleh karena itu, Anda dan para mitra Anda memutuskan untuk menarik minat investor. Mengingat Anda adalah seorang pakar analisis, para mitra pun meminta Anda untuk mempersiapkan sejumlah riset pasar. Anda memiliki data dari sumber terbuka (open source) terkait jenis-jenis usaha tempat makan di LA.

Tujuan

- Mengetahui jenis usaha yang paling banyak terdapat di LA.
- Mengetahui jumlah usaha pada setiap jalan.
- Melihat distribusi dari kursi yang dibutuhkan berdasarkan jenis usaha.

Import Library

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
```

```
import matplotlib.pyplot as plt
from scipy import stats as st
import numpy as np
from plotly import graph_objects as go
import plotly.express as px

import warnings
warnings.filterwarnings('ignore')
```

Pra Pemrosesan Data

```
In [2]: df = pd.read_csv('/datasets/rest_data_us.csv')
```

```
In [3]: df.head()
```

```
Out[3]:
```

	id	object_name	address	chain	object_type	number
0	11786	HABITAT COFFEE SHOP	3708 N EAGLE ROCK BLVD	False	Cafe	26
1	11787	REILLY'S	100 WORLD WAY # 120	False	Restaurant	9
2	11788	STREET CHURROS	6801 HOLLYWOOD BLVD # 253	False	Fast Food	20
3	11789	TRINITY ECHO PARK	1814 W SUNSET BLVD	False	Restaurant	22
4	11790	POLLEN	2100 ECHO PARK AVE	False	Restaurant	20

```
In [4]: df.shape
```

```
Out[4]: (9651, 6)
```

```
In [5]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9651 entries, 0 to 9650
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   id           9651 non-null   int64
1   object_name  9651 non-null   object
2   address      9651 non-null   object
3   chain        9648 non-null   object
4   object_type  9651 non-null   object
5   number       9651 non-null   int64
dtypes: int64(2), object(4)
memory usage: 452.5+ KB
```

```
In [6]: df.isna().sum()
```

```
Out[6]: id           0
object_name  0
address      0
chain        3
object_type  0
number       0
dtype: int64
```

```
In [7]: df.isna().sum()/len(df) * 100
```

```
Out[7]: id            0.000000
object_name  0.000000
address      0.000000
chain        0.031085
object_type  0.000000
number       0.000000
dtype: float64
```

```
In [8]: df = df.dropna().reset_index(drop=True)
```

```
In [9]: df[df.duplicated()]
```

```
Out[9]:
```

id	object_name	address	chain	object_type	number
----	-------------	---------	-------	-------------	--------

```
In [10]: df['object_type'].unique()
```

```
Out[10]: array(['Cafe', 'Restaurant', 'Fast Food', 'Bakery', 'Bar', 'Pizza'],
      dtype=object)
```

Terdapat nilai null pada dataset sebanyak 3 baris, nilai null tersebut dihapus karena hanya mempengaruhi 0,03% dari keseluruhan dataset. Tidak ada nilai duplikat pada dataset, sehingga dataset telah siap untuk diproses lebih lanjut.

Analisa Data

Proporsi Dari Berbagai Jenis Usaha

```
In [11]: df['object_type'].value_counts()
```

```
Out[11]: Restaurant    7253
Fast Food    1066
Cafe    435
Pizza    319
Bar    292
Bakery    283
Name: object_type, dtype: int64
```

```
In [12]: grouped_object = df.groupby('object_type').count().reset_index()
grouped_object
```

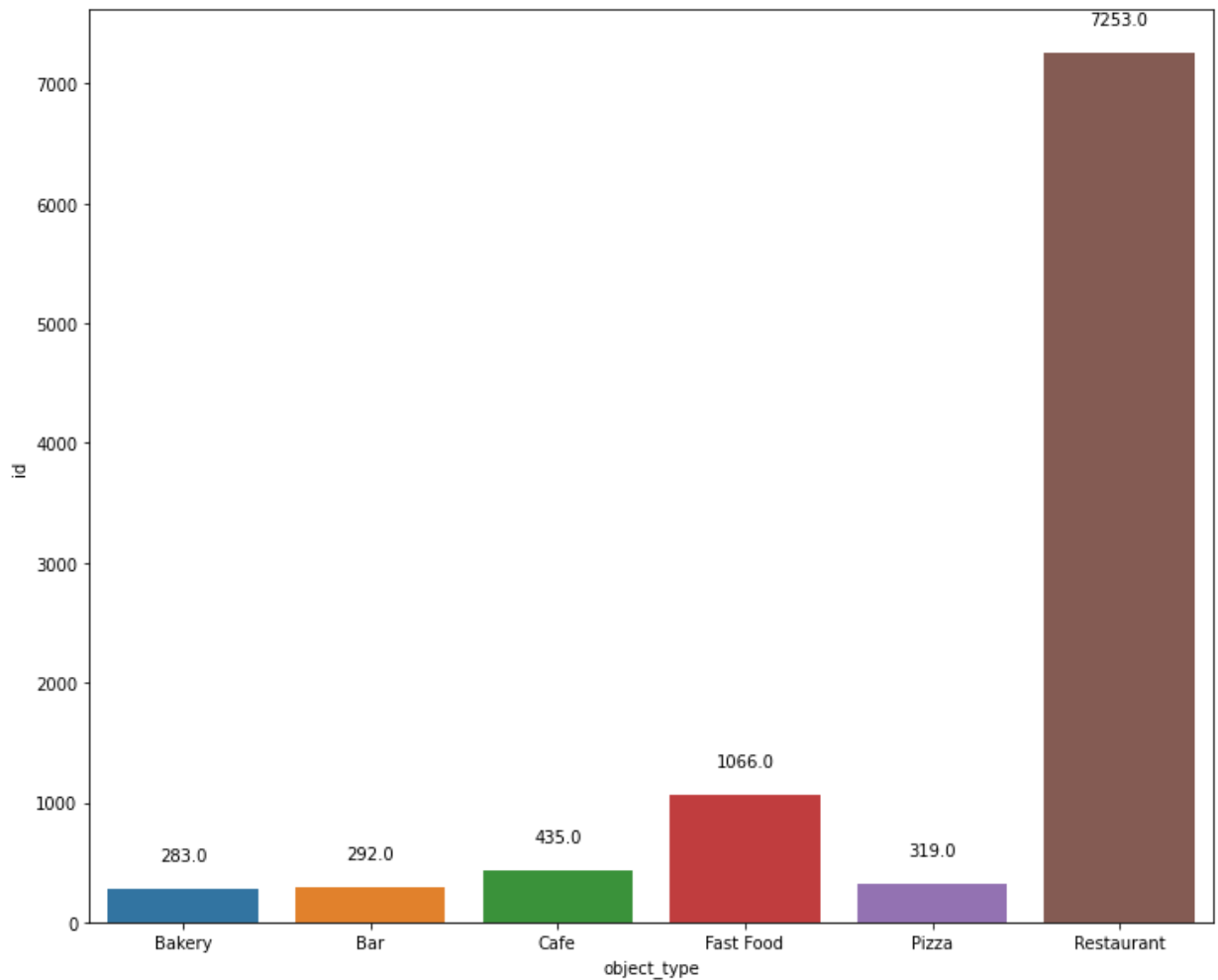
```
Out[12]:
```

	object_type	id	object_name	address	chain	number
0	Bakery	283	283	283	283	283
1	Bar	292	292	292	292	292
2	Cafe	435	435	435	435	435
3	Fast Food	1066	1066	1066	1066	1066
4	Pizza	319	319	319	319	319
5	Restaurant	7253	7253	7253	7253	7253

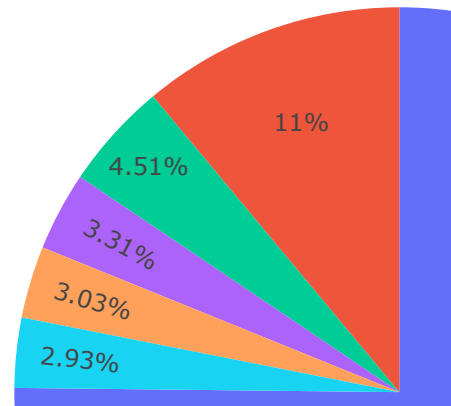
```
In [13]: plt.figure(figsize=(12,10))
sns.barplot(data=grouped_object, x='object_type', y='id')

for g in splot.patches:
    splot.annotate(format(g.get_height(), '.1f'),
                    (g.get_x() + g.get_width() / 2., g.get_height()),
                    ha = 'center', va = 'center',
                    xytext = (0, 20), rotation=0,
                    textcoords = 'offset points')

plt.show()
```



```
In [14]: fig = px.pie(grouped_object, values='id', names='object_type');
fig.show()
```



Dari diagram batang dan lingkaran diatas, terlihat jelas bahwa restaurant adalah usaha yang paling banyak proporsinya dalam dataset. Sebanyak 75,3% dari usaha yang ada merupakan restaurant, sisanya terbagi ke dalam bakery, fast food, bar, pizza, cafe. Bar dan bakery menjadi usaha dengan jumlah yang sedikit proporsinya.

Proporsi Dari Usaha Berbasis Waralaba dan Nonwaralaba

```
In [15]: grouped_chain = df.groupby('chain').count().reset_index()
         grouped_chain
```

```
Out[15]:
```

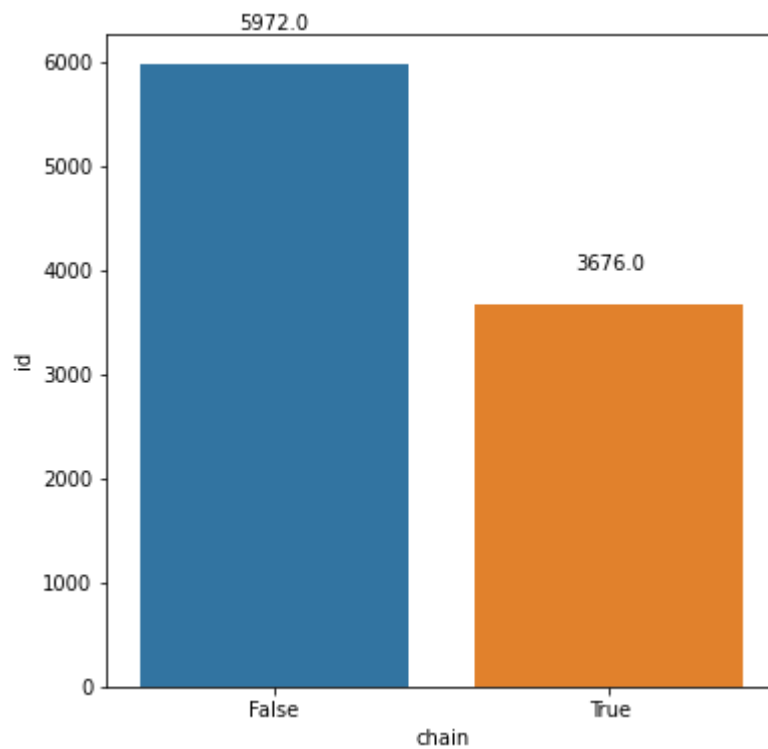
	chain	id	object_name	address	object_type	number
0	False	5972	5972	5972	5972	5972
1	True	3676	3676	3676	3676	3676

```
In [16]: plt.figure(figsize=(6,6))
         splot = sns.barplot(data=grouped_chain, x='chain', y='id')

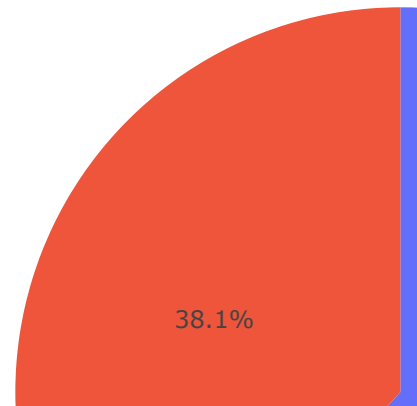
         for g in splot.patches:
             splot.annotate(format(g.get_height(), '.1f'),
                             (g.get_x() + g.get_width() / 2., g.get_height()),
                             ha = 'center', va = 'center',
```

```
xytext = (0, 20), rotation=0,  
textcoords = 'offset points')
```

```
plt.show()
```



```
In [17]: fig_chain = px.pie(grouped_chain, values='id', names='chain');  
fig_chain.show()
```



Sebanyak 5972 atau 61,9% usaha merupakan usaha non waralaba dan sisanya sebanyak 38,1% usaha merupakan usaha waralaba. Yang mana kebanyakan usaha di LA adalah usaha yang dimiliki oleh per seorangan, dan bukan merupakan franchise dari nama-nama besar yang sudah ada.

Jenis Usaha Yang Berbasis Waralaba

In [18]: df

Out[18]:

	id	object_name	address	chain	object_type	number
0	11786	HABITAT COFFEE SHOP	3708 N EAGLE ROCK BLVD	False	Cafe	26
1	11787	REILLY'S	100 WORLD WAY # 120	False	Restaurant	9
2	11788	STREET CHURROS	6801 HOLLYWOOD BLVD # 253	False	Fast Food	20
3	11789	TRINITY ECHO PARK	1814 W SUNSET BLVD	False	Restaurant	22
4	11790	POLLEN	2100 ECHO PARK AVE	False	Restaurant	20
...
9643	21432	HALL OF JUSTICE	217 W TEMPLE AVE	False	Restaurant	122
9644	21433	FIN-MELROSE	5750 MELROSE AVE	False	Restaurant	93
9645	21434	JUICY WINGZ	6741 HOLLYWOOD BLVD	True	Fast Food	15
9646	21435	MEDIDATE COFFEE	548 S SPRING ST STE 100	False	Cafe	6
9647	21436	CAFE SPROUTS	1300 S SAN PEDRO ST STE 111	True	Restaurant	19

9648 rows × 6 columns

```
In [19]: grouped_chain_object = pd.pivot_table(df, values='object_name',
        index='object_type', columns='chain',
        aggfunc='count').fillna(1)
grouped_chain_object = grouped_chain_object.reset_index(drop=False)
grouped_chain_object
```

Out[19]:

	chain	object_type	False	True
0	Bakery		1.0	283.0
1	Bar		215.0	77.0
2	Cafe		169.0	266.0
3	Fast Food		461.0	605.0
4	Pizza		166.0	153.0
5	Restaurant		4961.0	2292.0

```
In [20]: df[df['object_type']=='Bakery']['chain'].value_counts()
```

Out[20]: True 283
Name: chain, dtype: int64

```
In [21]: grouped_chain_object['ratio'] = (grouped_chain_object[True] / \
        (grouped_chain_object[True] + grouped_chain_object[False])
grouped_chain_object.sort_values(by='ratio')
```

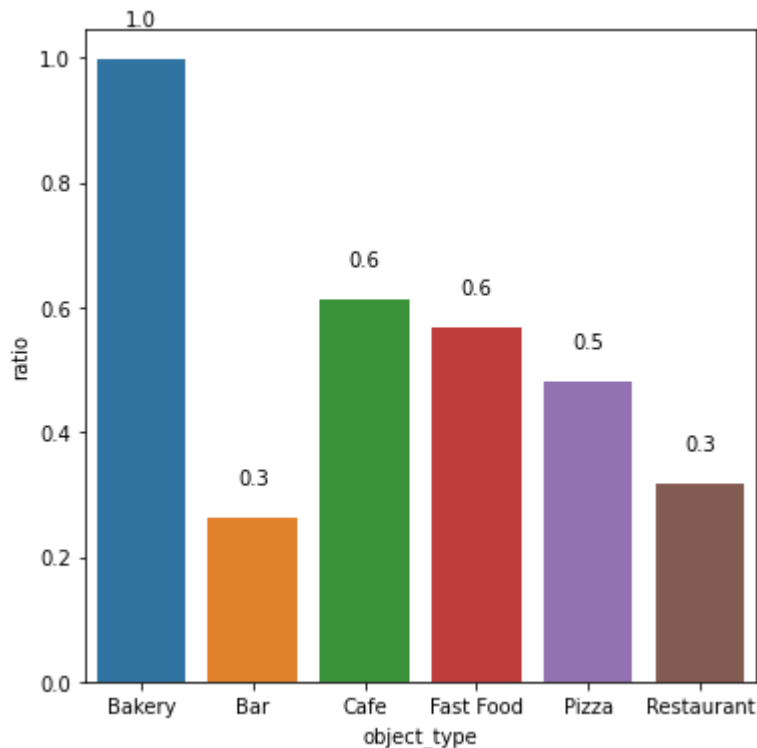

Out[21]:

chain	object_type	False	True	ratio
1	Bar	215.0	77.0	0.263699
5	Restaurant	4961.0	2292.0	0.316007
4	Pizza	166.0	153.0	0.479624
3	Fast Food	461.0	605.0	0.567542
2	Cafe	169.0	266.0	0.611494
0	Bakery	1.0	283.0	0.996479

```
In [22]: plt.figure(figsize=(6,6))
          sns.barplot(data=grouped_chain_object, x='object_type', y='ratio')

          for g in plot.patches:
              plot.annotate(format(g.get_height(), '.1f'),
                             (g.get_x() + g.get_width() / 2., g.get_height()),
                             ha = 'center', va = 'center',
                             xytext = (0, 20), rotation=0,
                             textcoords = 'offset points')

          plt.show()
```



Rasio tertinggi untuk perbandingan usaha waralaba dengan non waralaba adalah pada usaha bakery dan yang paling terkecil adalah usaha bar.

Yang Merupakan Ciri-Ciri Dari Jenis Usaha Berbasis Waralaba

```
In [23]: df_chain = df[df['chain']==True]
```

```
In [24]: df_chain['object_name'].value_counts().head()
```

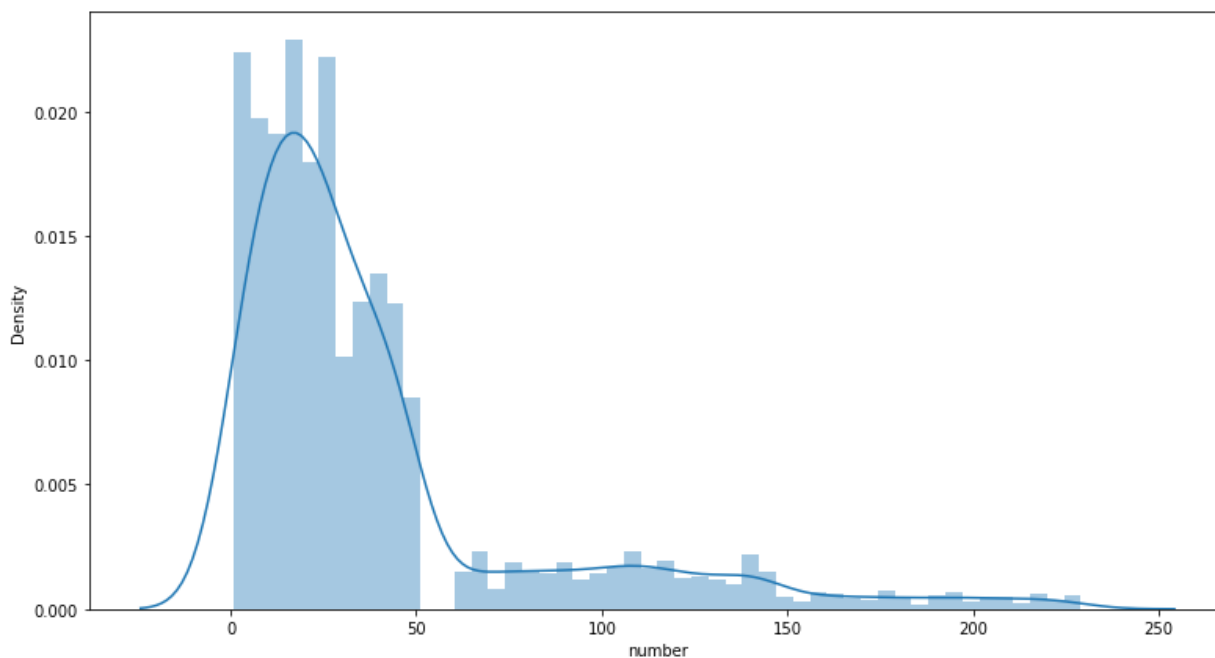
```
Out[24]: THE COFFEE BEAN & TEA LEAF    47
SUBWAY                               31
DOMINO'S PIZZA                       15
KENTUCKY FRIED CHICKEN               14
WABA GRILL                           14
Name: object_name, dtype: int64
```

```
In [25]: df_chain.head()
```

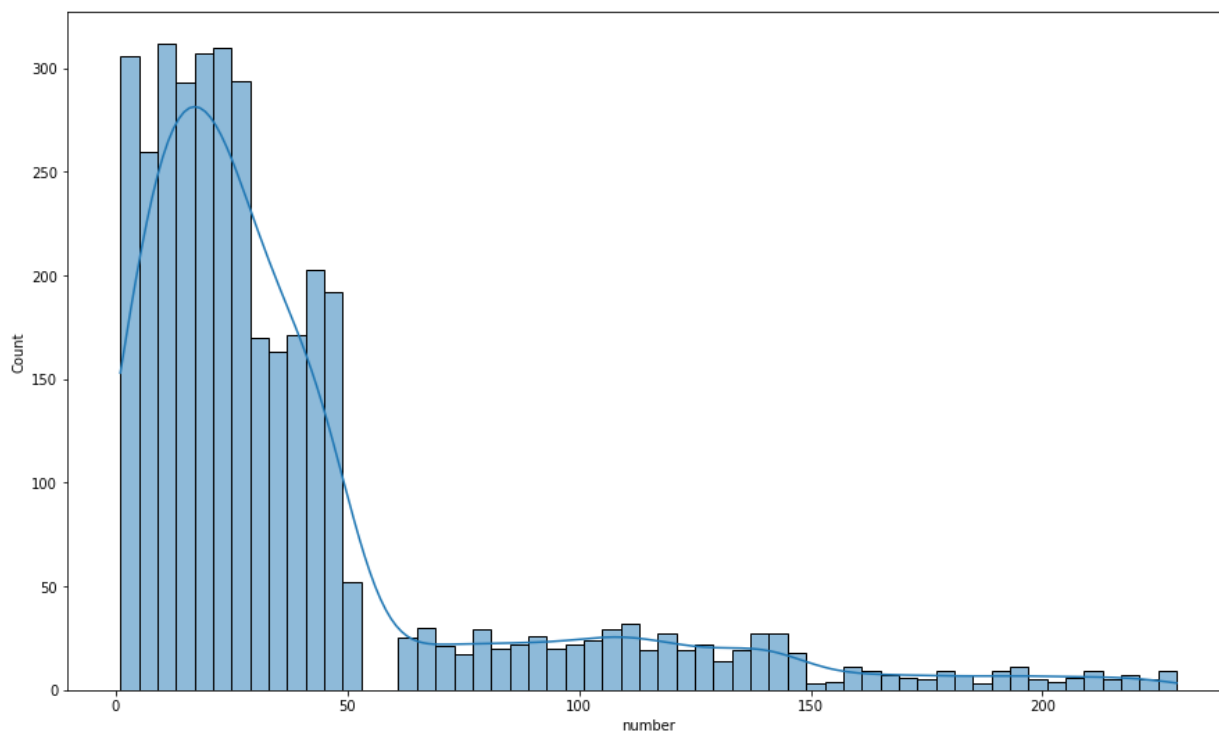
```
Out[25]:
```

	id	object_name	address	chain	object_type	number
8	11794	ABC DONUTS	3027 N SAN FERNANDO RD UNIT 103	True	Fast Food	1
10	11796	EL POLLO LOCO	5319 W SUNSET BLVD	True	Restaurant	38
11	11797	POONG NYUN BAKERY	928 S WESTERN AVE STE 109	True	Bakery	17
12	11798	EMC RESTAURANT GROUP LLC	3500 W 6TH ST STE 101	True	Restaurant	133
14	11800	CUSCATLECA BAKERY	2501 W SUNSET BLVD #A	True	Bakery	26

```
In [26]: plt.figure(figsize=(13, 7))
sns.distplot(df_chain['number'], kde=True);
```



```
In [27]: plt.figure(figsize=(15, 9))
sns.histplot(df_chain['number'], kde=True);
```



Usaha waralaba kebanyakan memiliki jumlah kursi di bawah 50 pada setiap tokonya. Untuk jumlah kursi di atas 50 hanya terdapat pada beberapa toko saja.

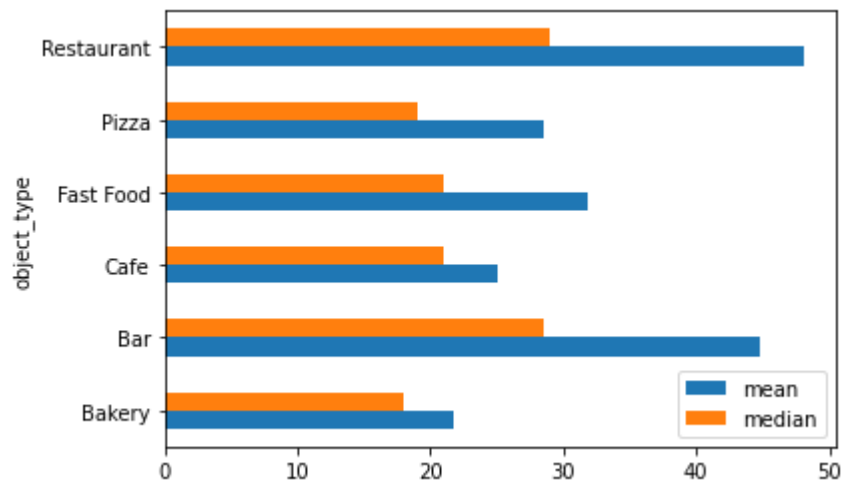
Rata-Rata Kursi Untuk Setiap Jenis Usaha

```
In [28]: df.groupby('object_type').agg(mean=('number', 'mean'), median=('number', 'median')).sort
```

```
Out[28]:
```

object_type	mean	median
Bakery	21.773852	18.0
Cafe	25.000000	21.0
Pizza	28.545455	19.0
Fast Food	31.837711	21.0
Bar	44.767123	28.5
Restaurant	48.048807	29.0

```
In [29]: df.groupby('object_type').agg(mean=('number', 'mean'), median=('number', 'median')).p
```



Usaha restaurant rata-rata memiliki jumlah kursi sebanyak 29 kursi diikuti dengan usaha bar dengan 28 kursi, usaha dengan jumlah rata-rata kursi yang paling sedikit adalah bakery.

Memisahkan Nama Alamat

In [30]: df

Out[30]:

	id	object_name	address	chain	object_type	number
0	11786	HABITAT COFFEE SHOP	3708 N EAGLE ROCK BLVD	False	Cafe	26
1	11787	REILLY'S	100 WORLD WAY # 120	False	Restaurant	9
2	11788	STREET CHURROS	6801 HOLLYWOOD BLVD # 253	False	Fast Food	20
3	11789	TRINITY ECHO PARK	1814 W SUNSET BLVD	False	Restaurant	22
4	11790	POLLEN	2100 ECHO PARK AVE	False	Restaurant	20
...
9643	21432	HALL OF JUSTICE	217 W TEMPLE AVE	False	Restaurant	122
9644	21433	FIN-MELROSE	5750 MELROSE AVE	False	Restaurant	93
9645	21434	JUICY WINGZ	6741 HOLLYWOOD BLVD	True	Fast Food	15
9646	21435	MEDIDATE COFFEE	548 S SPRING ST STE 100	False	Cafe	6
9647	21436	CAFE SPROUTS	1300 S SAN PEDRO ST STE 111	True	Restaurant	19

9648 rows × 6 columns

In [31]: df['address']

```
Out[31]: 0      3708 N EAGLE ROCK BLVD
1      100 WORLD WAY # 120
2      6801 HOLLYWOOD BLVD # 253
3      1814 W SUNSET BLVD
4      2100 ECHO PARK AVE
...
9643    217 W TEMPLE AVE
9644    5750 MELROSE AVE
9645    6741 HOLLYWOOD BLVD
9646    548 S SPRING ST STE 100
9647    1300 S SAN PEDRO ST STE 111
Name: address, Length: 9648, dtype: object
```

```
In [32]: df['address'].apply(lambda x: ' '.join(x.split(' ')[1:]))
```

```
Out[32]: 0      N EAGLE ROCK BLVD
1      WORLD WAY # 120
2      HOLLYWOOD BLVD # 253
3      W SUNSET BLVD
4      ECHO PARK AVE
...
9643    W TEMPLE AVE
9644    MELROSE AVE
9645    HOLLYWOOD BLVD
9646    S SPRING ST STE 100
9647    S SAN PEDRO ST STE 111
Name: address, Length: 9648, dtype: object
```

```
In [33]: df_address = df['address'].apply(lambda x: ' '.join(x.split(' ')[1:])).to_frame()
df_address
```

```
Out[33]:
```

	address
0	N EAGLE ROCK BLVD
1	WORLD WAY # 120
2	HOLLYWOOD BLVD # 253
3	W SUNSET BLVD
4	ECHO PARK AVE
...	...
9643	W TEMPLE AVE
9644	MELROSE AVE
9645	HOLLYWOOD BLVD
9646	S SPRING ST STE 100
9647	S SAN PEDRO ST STE 111

9648 rows × 1 columns

```
In [34]: df_address['count'] = 1
df_address
```

Out[34]:

	address	count
0	N EAGLE ROCK BLVD	1
1	WORLD WAY # 120	1
2	HOLLYWOOD BLVD # 253	1
3	W SUNSET BLVD	1
4	ECHO PARK AVE	1
...
9643	W TEMPLE AVE	1
9644	MELROSE AVE	1
9645	HOLLYWOOD BLVD	1
9646	S SPRING ST STE 100	1
9647	S SAN PEDRO ST STE 111	1

9648 rows × 2 columns

Dalam poin ini, akan dipisahkan beberapa string untuk dijadikan acuan dalam pengelompokkan. pada setiap baris juga diberikan nilai 1 untuk mempermudah menghitung usaha apa saja yang terdapat pada jalan yang sama.

Grafik Yang Menampilkan Sepuluh Nama Jalan Teratas

```
In [35]: top_address = df_address.groupby('address').count().sort_values(by='count', ascending=False)
top_address
```

Out[35]:

	address	count
0	W SUNSET BLVD	296
1	W PICO BLVD	288
2	HOLLYWOOD BLVD	167
3	WILSHIRE BLVD	161
4	S VERMONT AVE	148
...
3092	S BROADWAY # #A35	1
3093	S BROADWAY # 3	1
3094	S BROADWAY # 4CC/1	1
3095	S BROADWAY # A-B	1
3096	hollywood BLVD	1

3097 rows × 2 columns

```
In [36]: top_addresses = df_address.groupby('address').count().sort_values(by='count', ascending=False)
top_addresses
```

```
Out[36]:
```

	address	count
0	W SUNSET BLVD	296
1	W PICO BLVD	288
2	HOLLYWOOD BLVD	167
3	WILSHIRE BLVD	161
4	S VERMONT AVE	148
5	SANTA MONICA BLVD	146
6	W 3RD ST	145
7	BEVERLY BLVD	135
8	S FIGUEROA ST	134
9	MELROSE AVE	128

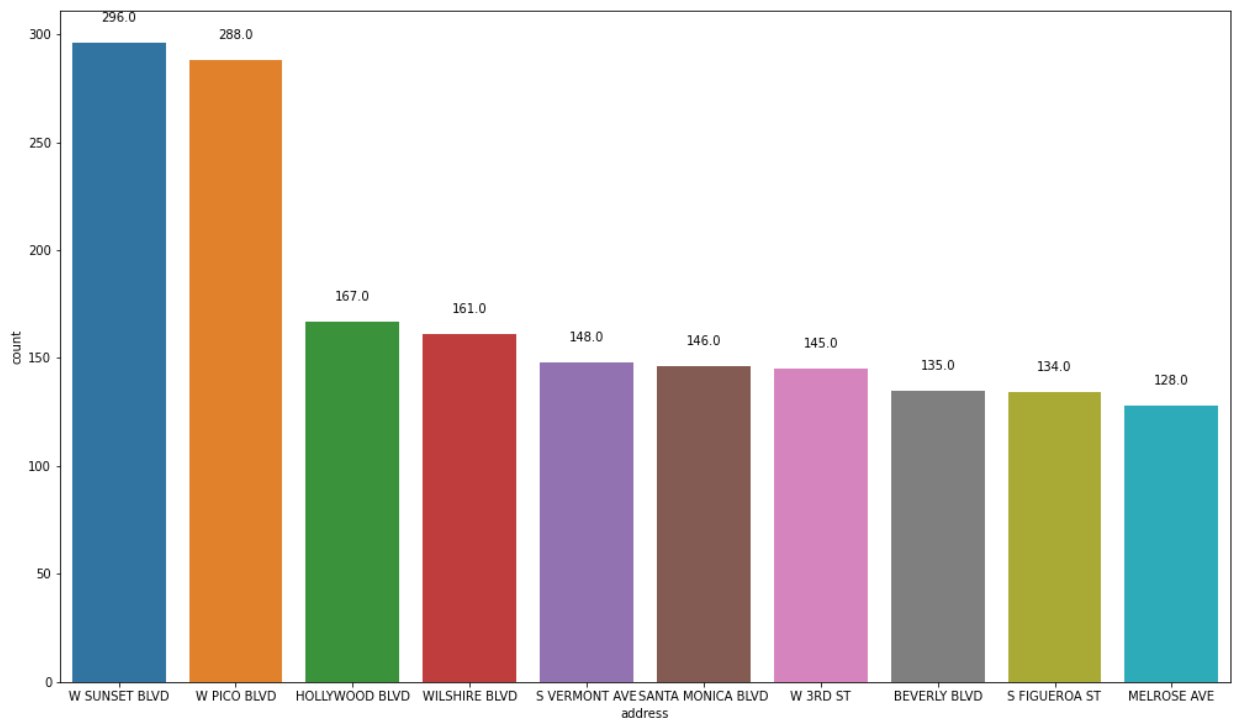
```
In [37]: top_addresses.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 2 columns):
#   Column  Non-Null Count  Dtype
---  -
0   address  10 non-null      object
1   count    10 non-null      int64
dtypes: int64(1), object(1)
memory usage: 288.0+ bytes
```

```
In [38]: plt.figure(figsize=(17,10))
plot = sns.barplot(data=top_addresses, x='address', y='count')

for g in plot.patches:
    plot.annotate(format(g.get_height(), '.1f'),
                  (g.get_x() + g.get_width() / 2., g.get_height()),
                  ha = 'center', va = 'center',
                  xytext = (0, 20), rotation=0,
                  textcoords = 'offset points')

plt.show()
```



Jalan W SUNSET BLVD adalah jalan dengan jumlah usaha terbanyak yaitu 296 usaha yang terdapat pada jalan tersebut. Jalan MELROSE AVE menjadi jalan yang paling sedikit dari 10 jalan dengan usaha terbanyak.

Jumlah Jalan Yang Memiliki Satu Restoran

```
In [39]: df_one = df_address.groupby('address').count().sort_values(by='count', ascending=False)
```

```
In [40]: df_one[df_one['count']==1].shape[0]
```

```
Out[40]: 2481
```

Sebanyak 2481 jalan hanya memiliki satu restoran.

Distribusi Jumlah Kursi

```
In [41]: top_address
```


Out[41]:

	address	count
0	W SUNSET BLVD	296
1	W PICO BLVD	288
2	HOLLYWOOD BLVD	167
3	WILSHIRE BLVD	161
4	S VERMONT AVE	148
...
3092	S BROADWAY # #A35	1
3093	S BROADWAY # 3	1
3094	S BROADWAY # 4CC/1	1
3095	S BROADWAY # A-B	1
3096	hollywood BLVD	1

3097 rows × 2 columns

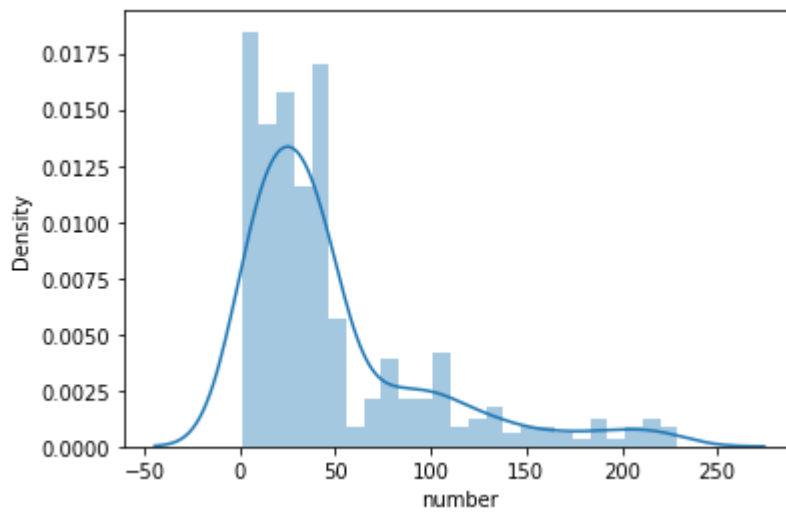
```
In [42]: top_street = df[df['address'].str.contains('W SUNSET B')].reset_index(drop=True)
top_street
```

Out[42]:

	id	object_name	address	chain	object_type	number
0	11789	TRINITY ECHO PARK	1814 W SUNSET BLVD	False	Restaurant	22
1	11796	EL POLLO LOCO	5319 W SUNSET BLVD	True	Restaurant	38
2	11800	CUSCATLECA BAKERY	2501 W SUNSET BLVD #A	True	Bakery	26
3	11817	DUNES RESTAURANT	5625 W SUNSET BLVD	True	Restaurant	118
4	11819	ONO HAWAIIAN BBQ	5539 W SUNSET BLVD	True	Restaurant	35
...
363	21284	PA ORD EXPRESS	5301 W SUNSET BLVD # 11	False	Fast Food	31
364	21309	BUTTER'S ICE CREAM	2925 W SUNSET BLVD	False	Restaurant	29
365	21313	REGULA GUY'S PIZZA	5173 W SUNSET BLVD	False	Pizza	20
366	21348	7-ELEVEN STORE #18778D	7600 W SUNSET BLVD	False	Restaurant	1
367	21368	HYDE LLC	8029 W SUNSET BLVD	False	Restaurant	42

368 rows × 6 columns

```
In [43]: sns.distplot(top_street['number']);
```



Pada jalan W SUNSET BLVD, kebanyakan usaha yang terdapat pada jalan itu memiliki jumlah kursi di antara 1 sampai 50, sedangkan hanya beberapa usaha saja yang memiliki jumlah kursi diatas 50. Hal itu terlihat pada grafik diatas, dimana setelah angka 50 grafik terlihat cukup melandai.

Kesimpulan

- Usaha yang paling banyak mendominasi adalah restaurant dengan 75,3% dari keseluruhan proporsi usaha
- Sebanyak 61,9% usaha merupakan usaha non waralaba dengan jumlah usaha 5972
- Usaha restaurant memiliki rata-rata jumlah kursi sebesar 29 buah diikuti oleh bar dengan 28 buah, yang paling sedikit adalah usaha bakery dengan 18 kursi
- Jalan W SUNSET BLVD menjadi jalan dengan jumlah usaha terbanyak sebesar 296 usaha
- Link Google Slide:
<https://drive.google.com/file/d/1M5idJs1WMYCLAfZrW3ClEjaceOvvkJIH/view?usp=sharing>