

project_10_test

February 12, 2025

1 Projeto 10

1.0.1 Descrição do Projeto

Contexto: Você decidiu abrir uma pequena cafeteria com garçons robôs em Los Angeles. É um projeto bem promissor, mas caro, então você e seus parceiros decidem tentar atrair investidores. Eles estão interessados nas condições atuais do mercado e querem saber se você conseguirá manter seu sucesso quando a novidade de garçons robôs esmorecer. Você é um guru de análise, então seus parceiros te pediram para preparar uma pesquisa do mercado. Você tem dados de código aberto sobre restaurantes em LA.

1.0.2 Índice

1. Carregamento dos dados e preparo para análise
2. Análise de dados
 - Proporções de vários tipos de estabelecimentos
 - Proporções de estabelecimentos de rede e não rede
 - Tipos de estabelecimento típico para redes
 - O que caracteriza redes: muitos estabelecimentos com um pequeno número de assentos ou poucos estabelecimentos com muitos assentos?
 - Número médio de assentos para cada tipo de estabelecimento
 - Investigando locais mais concorridos
 - Top 10 ruas com o maior número de restaurantes
 - Número de ruas que têm apenas um restaurante
 - Distribuição de número de assentos para as ruas com muitos restaurantes
 - Proporção tipo de estabelecimentos nas ruas mais famosas
3. Conclusão

Passo 1. Carregamento dos dados e preparo para a análise

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
from matplotlib import pyplot as plt
from plotly import graph_objects as go
import plotly.express as px
import warnings
warnings.filterwarnings("ignore")
```

```
sns.set_context('notebook', font_scale=1, rc={"lines.linewidth": 2.5})
```

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

```
[3]: 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
```

- Colunas com formato de dados correto
- Colunas no formato snake case
- 3 dados ausentes na coluna chain

```
[4]: df.head()
```

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

      number
0         26
1          9
2         20
3         22
4         20
```

```
[5]: # checking duplicates
```

```
print(df.duplicated().sum())
print(df.duplicated(subset=['id']).sum())
print(df.duplicated(subset=['object_name', 'address']).sum())
```

```
0
0
0
```

```
[6]: # checking null values
df[df['chain'].isna()]
```

```
[6]:
```

	id	object_name	address	chain	object_type	\
	7408	19194	TAQUERIA LOS 3 CARNALES	5000 E WHITTIER BLVD	NaN	Restaurant
	7523	19309	JAMMIN JIMMY'S PIZZA	1641 FIRESTONE BLVD	NaN	Pizza
	8648	20434	THE LEXINGTON THEATER	129 E 3RD ST	NaN	Restaurant


```

number
7408    14
7523     1
8648    35

```

```
[7]: chain_nan = ["TAQUERIA LOS 3 CARNALES",
                  "JAMMIN JIMMY'S PIZZA", "THE LEXINGTON THEATER"]
df[df['object_name'].isin(chain_nan)]
```

```
[7]:
```

	id	object_name	address	chain	object_type	\
	7408	19194	TAQUERIA LOS 3 CARNALES	5000 E WHITTIER BLVD	NaN	Restaurant
	7523	19309	JAMMIN JIMMY'S PIZZA	1641 FIRESTONE BLVD	NaN	Pizza
	8648	20434	THE LEXINGTON THEATER	129 E 3RD ST	NaN	Restaurant


```

number
7408    14
7523     1
8648    35

```

Para ser considerado uma rede um restaurante deve possuir 2 estabelecimentos ou mais, nesse caso vimos que os restaurantes com valor ausente na coluna chain só possui um restaurante, portanto o valor ausente vai ser substituído por False.

```
[8]: df['chain'].fillna(False, inplace=True)
df[df['chain'].isna()]
```

```
[8]: Empty DataFrame
Columns: [id, object_name, address, chain, object_type, number]
Index: []
```

```
[9]: df.describe(include='object')
```

```
[9]:
```

	object_name	address	object_type
count	9651	9651	9651
unique	8672	8517	6
top	THE COFFEE BEAN & TEA LEAF	3607 TROUSDALE PKWY	Restaurant
freq	47	11	7255

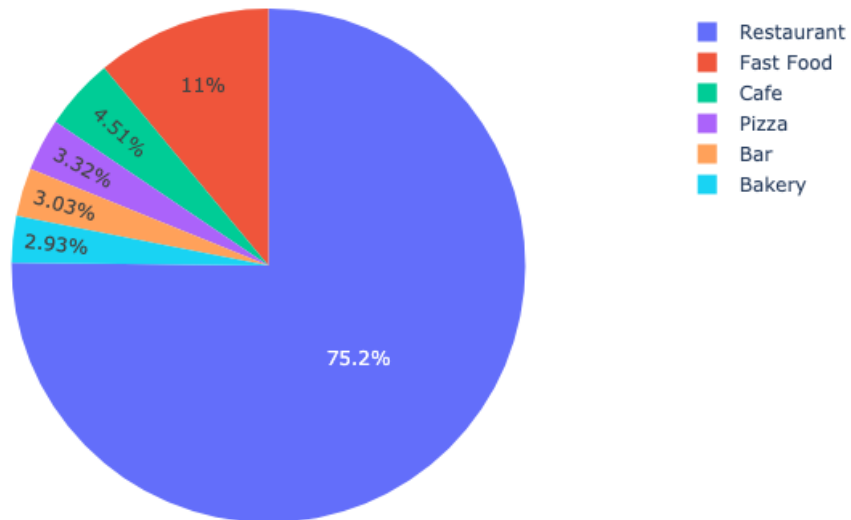
Passo 2. Análise de dados

Proporções de vários tipos de estabelecimentos

```
[10]: # count object_type
type_count = df['object_type'].value_counts()

# plot
fig = go.Figure(
    data=[go.Pie(labels=type_count.index, values=type_count.values)])
fig.update_layout(title_text='Proporção dos Tipos de Estabelecimentos',
                    title_x=0.5)
fig.show('png')
```

Proporção dos Tipos de Estabelecimentos



Maioria esmagadora de restaurantes

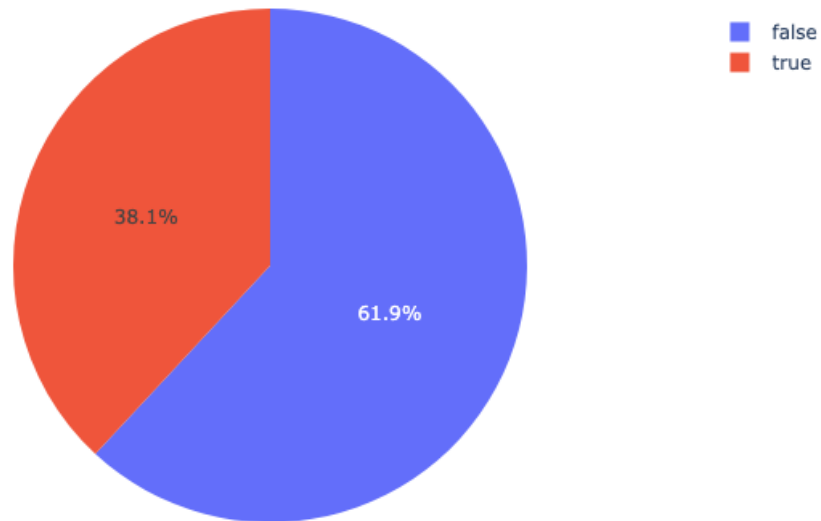
Proporções de estabelecimentos de rede e não rede

```
[11]: # chain counts
chain_counts = df['chain'].value_counts()

# plot
fig = go.Figure(
    data=[go.Pie(labels=chain_counts.index, values=chain_counts.values)])
fig.update_layout(
    title_text='Proporção de Estabelecimentos que são de Rede (Chain)',
```

```
title_x=0.5)
fig.show('png')
```

Proporção de Estabelecimentos que são de Rede (Chain)



Maioria não é rede, 61.9%

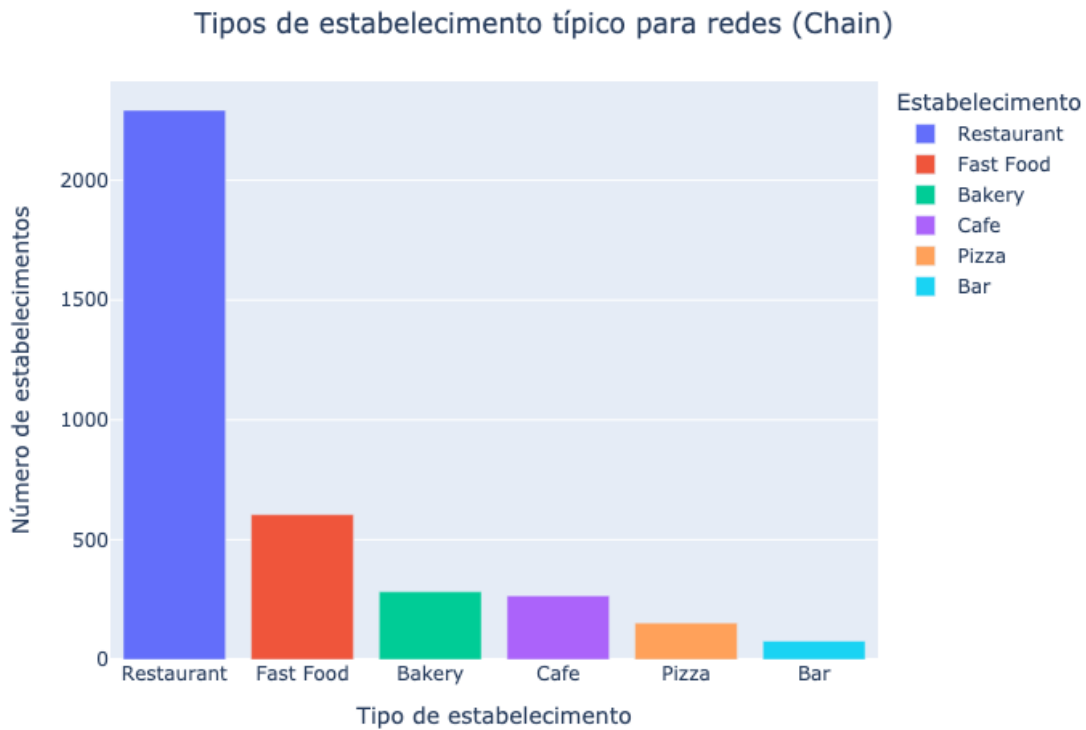
Tipos de estabelecimento típico para redes

```
[12]: # filter
df_chain_true = df[df['chain'] == True]

# group by type
df_group_chain_obtype = df_chain_true.groupby(
    'object_type')['object_type'].count().sort_values(ascending=False)

# plot
fig = px.bar(df_group_chain_obtype, x=df_group_chain_obtype.index,
             y=df_group_chain_obtype.values, color=df_group_chain_obtype.index)
fig.update_layout(
    title="Tipos de estabelecimento típico para redes (Chain)",
    xaxis_title="Tipo de estabelecimento",
    yaxis_title="Número de estabelecimentos",
    title_x=0.5, # Centraliza o título
    legend_title="Estabelecimento")
```

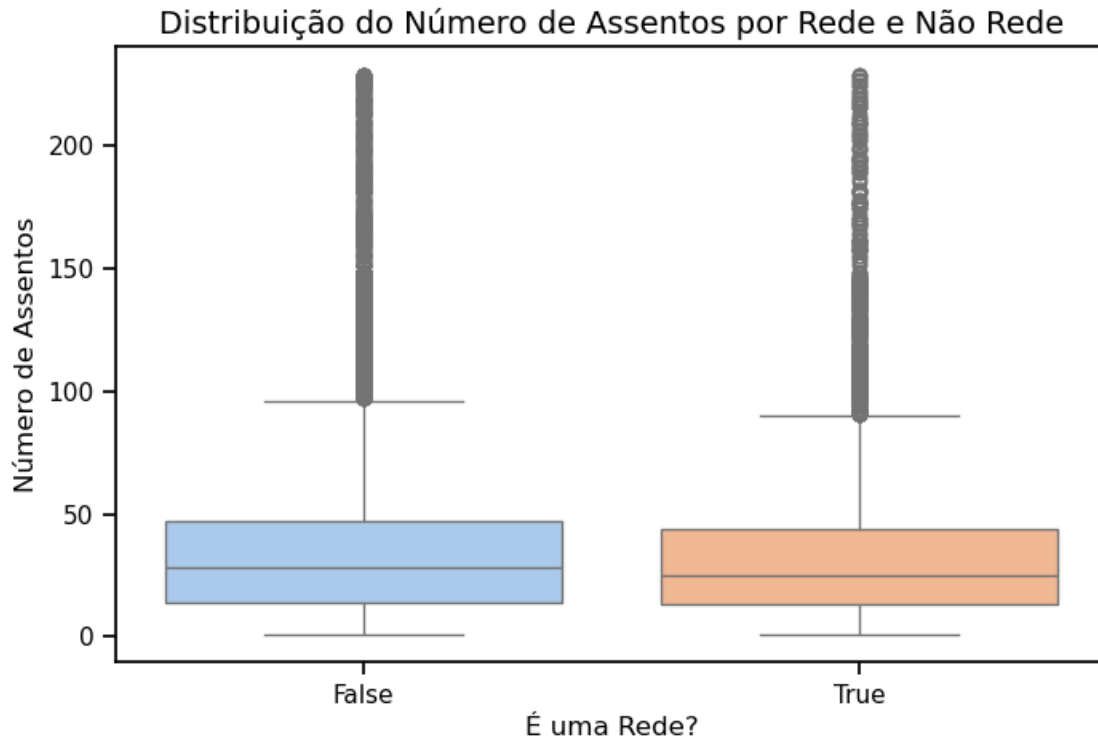
```
)  
  
fig.show('png')
```



Maioria esmagadora de restaurante como estabelecimento típicos de redes, mas também temos o fast food com bons números

O que caracteriza redes: muitos estabelecimentos com um pequeno número de assentos ou poucos estabelecimentos com muitos assentos?

```
[13]: # Boxplot number of seats, chain and no chain  
  
plt.figure(figsize=(8, 5))  
sns.boxplot(data=df, x="chain", y="number", palette='pastel')  
  
plt.title("Distribuição do Número de Assentos por Rede e Não Rede", fontsize=14)  
plt.xlabel("É uma Rede?", fontsize=12)  
plt.ylabel("Número de Assentos", fontsize=12)  
  
plt.show()
```



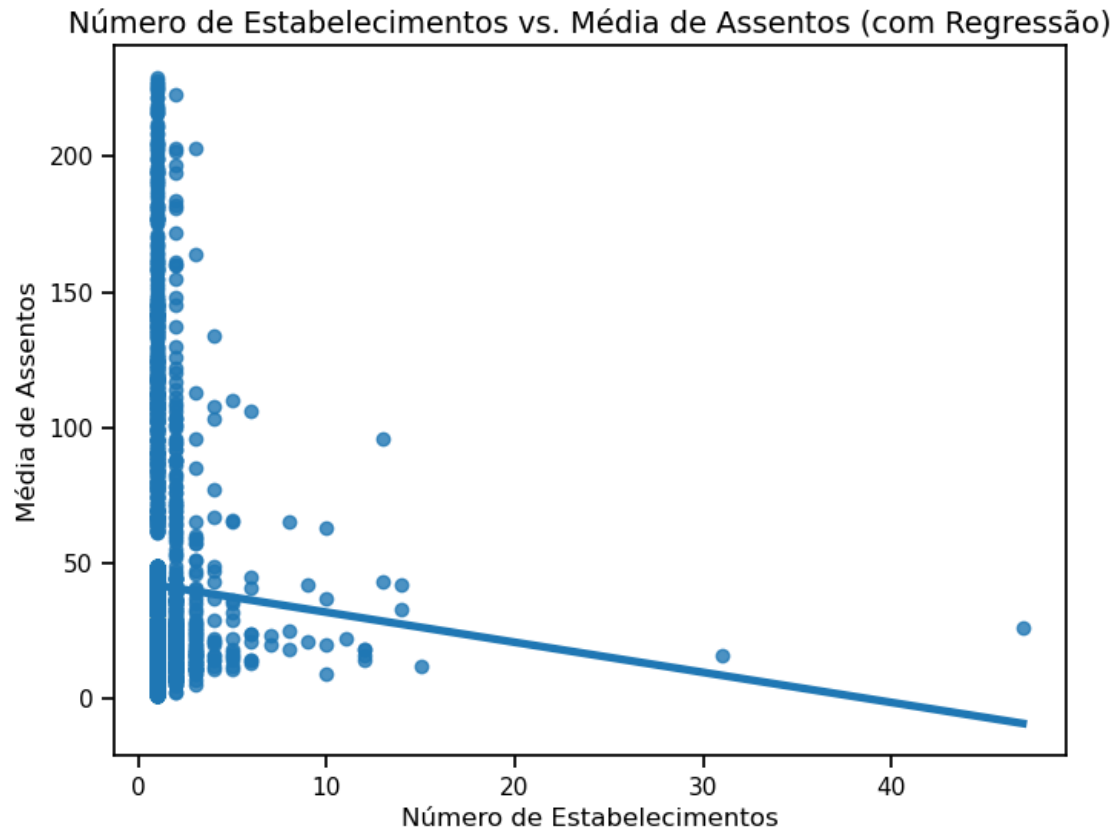
Somente por número de assentos não se caracteriza como rede ou não, boxplot tem números parecidos

```
[14]: # count number of seats per chain
df_chain_seats = df_chain_true.groupby('object_name').
    ↪agg(total_estabelecimentos=(
        "id", "count"), media_assentos=("number", "mean")).reset_index()
df_chain_seats['media_assentos'] = df_chain_seats['media_assentos'].astype(int)

# scatter plot
plt.figure(figsize=(8, 6))
sns.regplot(data=df_chain_seats, x="total_estabelecimentos",
            y="media_assentos", scatter=True, ci=None)

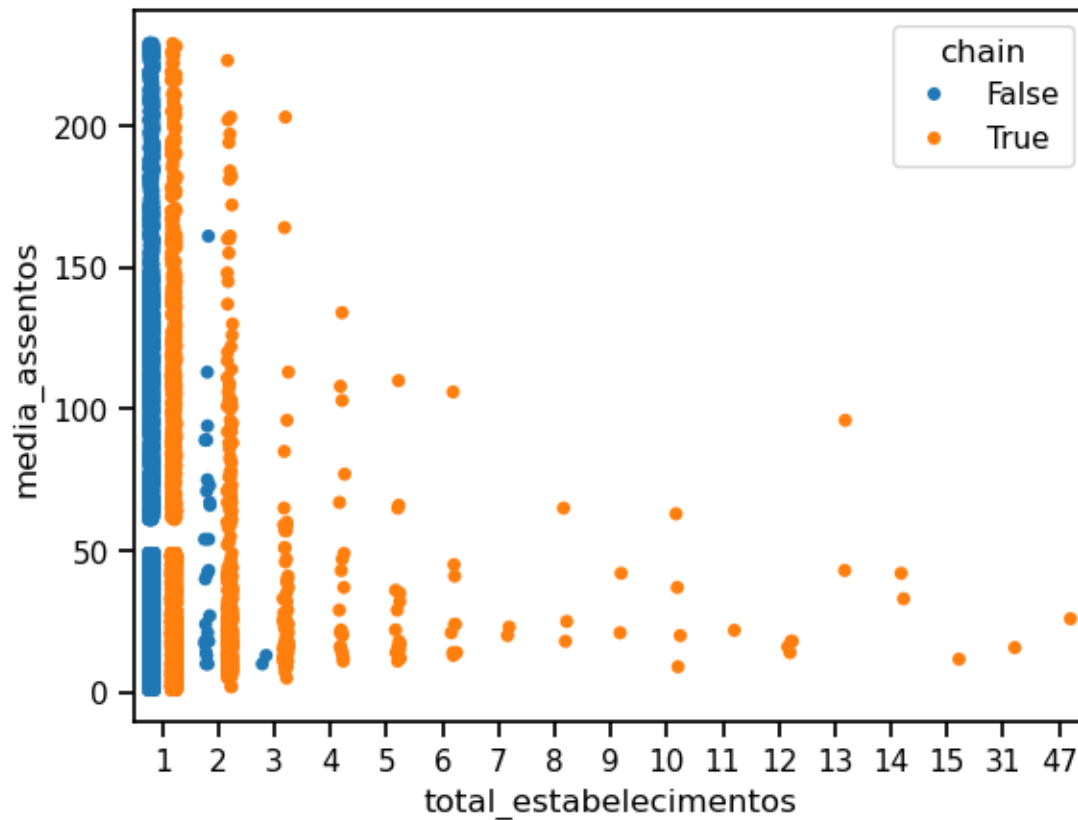
plt.title(
    "Número de Estabelecimentos vs. Média de Assentos (com Regressão)",
    ↪fontsize=14)
plt.xlabel("Número de Estabelecimentos", fontsize=12)
plt.ylabel("Média de Assentos", fontsize=12)

plt.show()
```



```
[15]: # stripplot
df_grouped = df.groupby(['chain', 'object_name']).agg(total_estabelecimentos=(
    "id", "count"), media_assentos=("number", "mean")).reset_index()
df_grouped['media_assentos'] = df_grouped['media_assentos'].astype(int)

sns.stripplot(x="total_estabelecimentos", y="media_assentos",
              data=df_grouped, hue='chain', dodge=True, jitter=True)
plt.show()
```

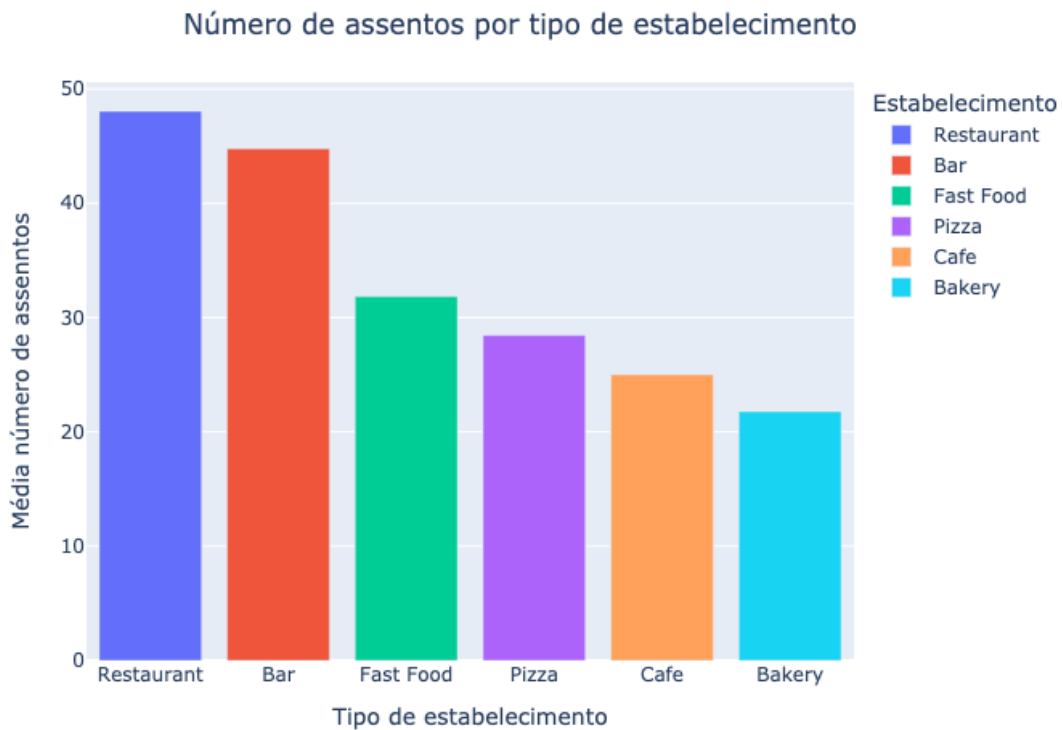
Segundo os gráficos de dispersão as redes se caracterizam por poucos estabelecimentos com média de muitos assentos

Número médio de assentos para cada tipo de estabelecimento

```
[16]: # group by
df_seats_per_obtype = df.groupby('object_type')['number'].mean().reset_index()
df_seats_per_obtype.sort_values(by='number', ascending=False, inplace=True)

# plot
fig = px.bar(df_seats_per_obtype, x='object_type',
             y='number', color='object_type')
fig.update_layout(
    title="Número de assentos por tipo de estabelecimento",
    xaxis_title="Tipo de estabelecimento",
    yaxis_title="Média número de assentos",
    title_x=0.5,
    legend_title="Estabelecimento"
)

fig.show('png')
```



Restaurante e bar são os estabelecimentos com maiores média de número de assentos

Investigando locais mais concorridos

```
[17]: # saving just street name
df['street_name'] = df['address'].str.replace(
    r'[~a-zA-ZÃ-ÿ\s]', '', regex=True)

# regex=True: regular expression
# r'[~a-zA-ZÃ-ÿ\s]': Removes everything other than:
# Uppercase letters (A-Z)
# Lowercase letters (a-z)
# Accented letters (Ã-ÿ)
# Spaces (\s)

df.head()
```

```
[17]:
```

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

	number	street_name
0	26	N EAGLE ROCK BLVD
1	9	WORLD WAY
2	20	HOLLYWOOD BLVD
3	22	W SUNSET BLVD
4	20	ECHO PARK AVE

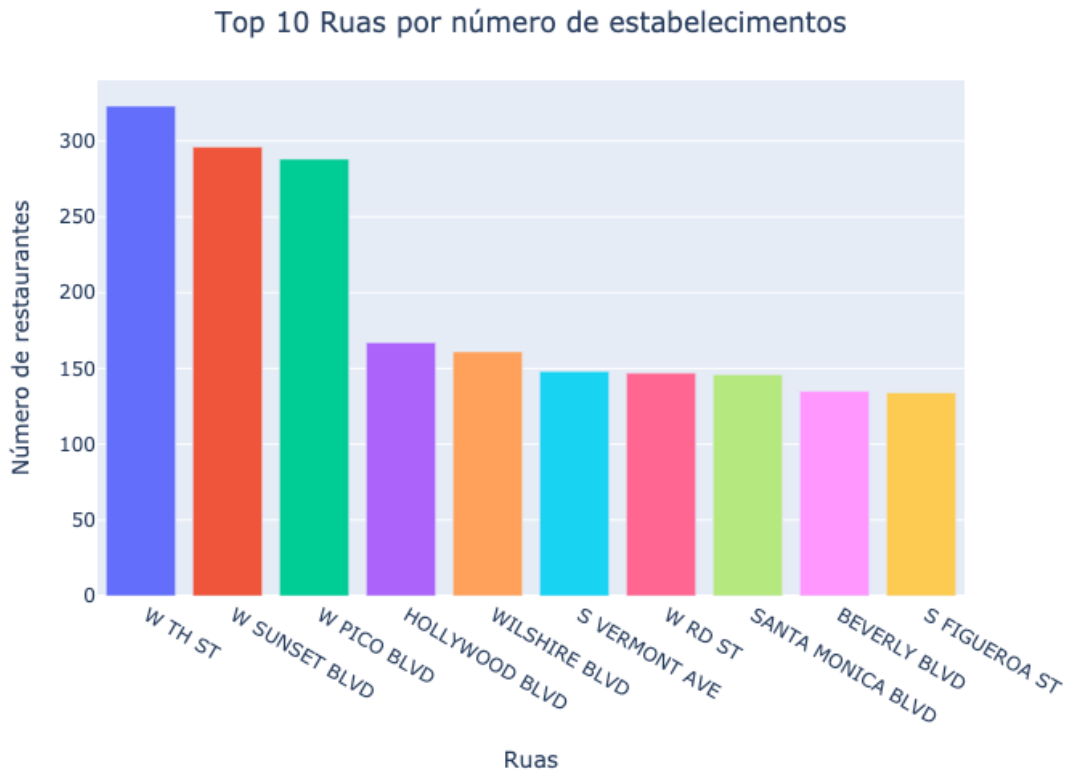
Top 10 ruas com o maior número de restaurantes

```
[18]: restaurants_by_streets = df.groupby('street_name')['id'].count().reset_index()

# top10
top10_streets = restaurants_by_streets.sort_values(
    by='id', ascending=False).head(10)

# plt
fig = px.bar(top10_streets, x='street_name',
             y='id', color='street_name')
fig.update_layout(
    title="Top 10 Ruas por número de estabelecimentos",
    xaxis_title="Ruas",
    yaxis_title="Número de restaurantes",
    title_x=0.5,
    showlegend=False
)

fig.show('png')
```



W TH ST, W SUNSET BLVD, W PICO BLVD são as três ruas com maior número de estabelecimentos

Número de ruas que têm apenas um restaurante.

```
[19]: len(restaurants_by_streets[restaurants_by_streets['id'] == 1])
```

```
[19]: 1096
```

1096 ruas com apenas um restaurante

Distribuição de número de assentos para as ruas com muitos restaurantes.

```
[20]: list_top_streets = top10_streets['street_name']

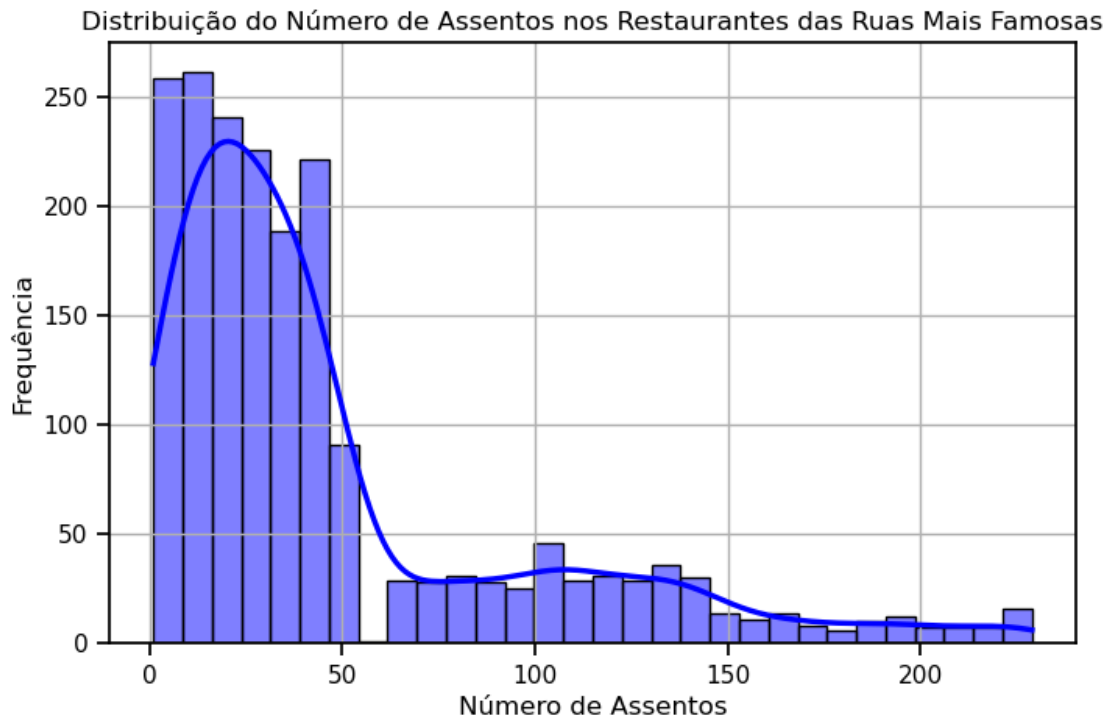
df_filtered_top_streets = df[df['street_name'].isin(list_top_streets)]

# plot
plt.figure(figsize=(8, 5))
sns.histplot(df_filtered_top_streets["number"],
             kde=True, bins=30, color='blue')

# Personalizar
```

```
plt.title("Distribuição do Número de Assentos nos Restaurantes das Ruas Mais_
↪Famosas", fontsize=12)
plt.xlabel("Número de Assentos", fontsize=12)
plt.ylabel("Frequência", fontsize=12)
plt.grid(True)

# Exibir gráfico
plt.show()
```



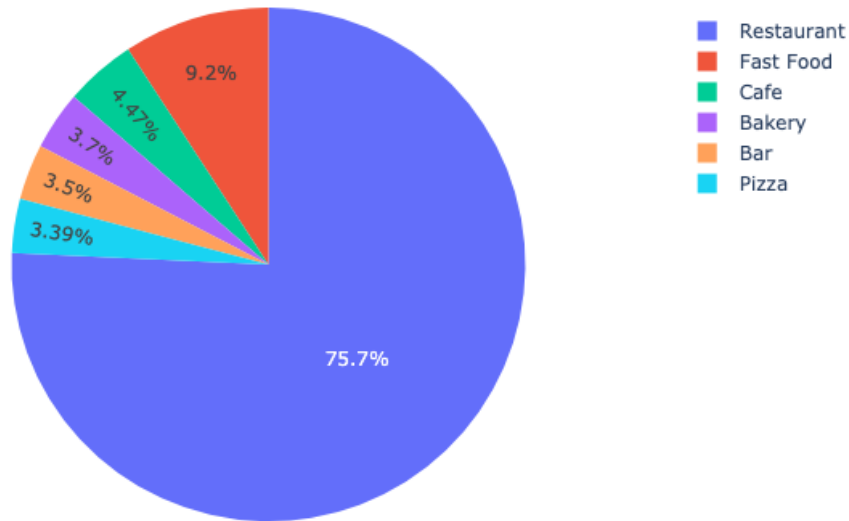
Nas ruas mais requisitadas para esse tipo de serviço, a uma tendência dos restaurantes em ter menos 50 assentos

Proporção tipo de estabelecimentos nas ruas mais famosas

```
[21]: type_count = df_filtered_top_streets['object_type'].value_counts()

# plot
fig = go.Figure(
    data=[go.Pie(labels=type_count.index, values=type_count.values)])
fig.update_layout(title_text='Proporção dos Tipos de Estabelecimentos nas Ruas_
↪mais Famosas',
                    title_x=0.5)
fig.show('png')
```

Proporção dos Tipos de Estabelecimentos nas Ruas mais Famosas



1.0.3 Conclusão: Conclusão geral e recomendações

Analisando os gráficos nota-se que há certas tendências como: * Maioria esmagadora dos estabelecimentos são restaurantes tanto nas ruas mais famosas como no geral * Há mais restaurantes de redes (60% contra 40%) * Segundo os gráficos de dispersão as redes se caracterizam por poucos estabelecimentos com média de muitos assentos * Restaurante e bar são os estabelecimentos com maior média de número de assentos (50) * Nas ruas mais requisitadas para esse tipo de serviço, a uma tendência dos restaurantes em ter menos de 50 assentos

Como há uma concorrência muito forte para o tipo de estabelecimento restaurante, é super recomendável abrir um outro tipo de estabelecimento, como um bar ou café, sobre o número de assentos dependeria do local de abertura, se fosse o caso de abrir em um local famoso onde já existe bastante estabelecimentos, recomendaria um lugar com poucos assentos por conta do preço do aluguel que provavelmente deve ser mais caro e também poderia dar um ar de ‘exclusividade’ não só pelo o número limitado de assentos mais por conta da novidade de garçons rôbos. Já se abertura do estabelecimento não fosse em um lugar tão requisitados, recomendaria um número alto de assentos para um retorno do investimento mais rápido. Dependendo do sucesso do novo negócio, recomendaria investir em uma rede mas seguindo a mesma mentalidade inicial de (lugares famosos, assentos exclusivos / lugares comuns, muitos assentos)

Report em pdf e apresentação html desse projeto: <[Github](#)>