In [4]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df=pd.read_csv("food_dely.csv")
df.head(40)
```

Out[4]:

	res_id	name	establishment	url	address	city	city_id	local
0	3400299	Bikanervala	Quick Bites	https://www.zomato.com/agra/bikanervala-khanda...	Kalyani Point, Near Tulsi Cinema, Bypass Road,...	Agra	34	Khanc
1	3400005	Mama Chicken Mama Franky House	Quick Bites	https://www.zomato.com/agra/mama-chicken-mama-...	Main Market, Sadar Bazaar, Agra Cantt, Agra	Agra	34	Agra Ca
2	3401013	Bhagat Halwai	Quick Bites	https://www.zomato.com/agra/bhagat-halwai-2-sh...	62/1, Near Easy Day, West Shivaji Nagar, Goalp...	Agra	34	Shahg
3	3400290	Bhagat Halwai	Quick Bites	https://www.zomato.com/agra/bhagat-halwai-civi...	Near Anjana Cinema, Nehru Nagar, Civil Lines, ...	Agra	34	Civil Li
4	3401744	The Salt Cafe Kitchen & Bar	Casual Dining	https://www.zomato.com/agra/the-salt-cafe-kitc...	1C,3rd Floor, Fatehabad Road, Tajganj, Agra	Agra	34	Tajg
5	3400275	Dominos Pizza	Quick Bites	https://www.zomato.com/agra/dominos-pizza-civi...	114/23 G, Deep Shikha Complex, Sanjay Place, C...	Agra	34	Civil Li
6	3400296	Honeydew Restaurant	Quick Bites	https://www.zomato.com/agra/honeydew-restauran...	Opposite Soami Bagh Temple, Dayal Bagh, Agra	Agra	34	Da Ba

	res_id	name	establishment	url	address	city	city_id	local
7	3400368	Dominos Pizza	Quick Bites	https://www.zomato.com/agra/dominos-pizza-sika...	Plot C-1/6, Sector 13, Sikandra, Agra	Agra	34	Sikan
8	3401284	Cake House	Bakery	https://www.zomato.com/agra/cake-house-2-civil...	23/301, Wazirpura Rd, Judge Compound Chowraha,...	Agra	34	Civil Li
9	3400838	Sugar N Thyme	Café	https://www.zomato.com/agra/sugar-n-thyme-tajg...	1374 K/1375 K, Ground floor, Dinesh Nagar, Fat...	Agra	34	Tajg
10	3400868	Hichkee	Casual Dining	https://www.zomato.com/agra/hichkee-civil-line...	6/27/1A, Hotel Alleviate, Yamuna Kinara Road, ...	Agra	34	Civil Li
11	3400425	Bikanervala	Casual Dining	https://www.zomato.com/agra/bikanervala-tajgan...	Opposite ITC Mughal, Fatehbad Road, Tajganj, Agra	Agra	34	Tajg
12	3400105	Pizza Hut	Casual Dining	https://www.zomato.com/agra/pizza-hut-tajganj?...	8, Handicraft Nagar, Fatehabad Road, Tajganj, ...	Agra	34	Tajg
13	3400457	Yadav Dhaba Purana	Dhaba	https://www.zomato.com/agra/yadav-dhaba-purana...	1/48, Delhi Gate, Station Road, Raja Mandi, Ci...	Agra	34	Civil Li
14	3400282	Aahar	Casual Dining	https://www.zomato.com/agra/aahar-civil-lines?...	26/118, Opposite Income Tax Building, Sanjay P...	Agra	34	Civil Li
15	3400611	Dominos Pizza	Quick Bites	https://www.zomato.com/agra/dominos-	Ground & First Floor, City Mall, Balkeshwar Co...	Agra	34	Kar Na

	res_id	name	establishment	url	address	city	city_id	local
				pizza-kaml...				
16	3400017	Pinch Of Spice	Casual Dining	https://www.zomato.com/agra/pinch-of-spice-civ...	23/453, Opposite Sanjay Cinema, Wazipura Road,...	Agra	34	Civil Li
17	3400279	Mc Donalds	Quick Bites	https://www.zomato.com/agra/mc-donalds-civil-l...	4-5, Ashok Cosmos Mall, Sanjay Place, Civil Li...	Agra	34	Civil Li
18	3400010	GMB - Gopika Sweets & Restaurant	Casual Dining	https://www.zomato.com/agra/gmb-gopika-sweets-...	Crystal Tower, Fatehabad Road, Tajganj, Agra	Agra	34	Tajg
19	3400712	The Greener Side	Quick Bites	https://www.zomato.com/agra/the-greener-side-c...	Delhi Gate Rd	Agra	34	Civil Li
20	3400297	Bhagat Halwai	Quick Bites	https://www.zomato.com/agra/bhagat-halwai-daya...	Heera Bagh, Khel Gaon, Dayal Bagh, Agra	Agra	34	Da Ba
21	3400370	Pizza Hut	Casual Dining	https://www.zomato.com/agra/pizza-hut-sikandra...	Hotel Adhiraj Palace, Opposite Kargil Petrol P...	Agra	34	Sikan
22	3400018	Pinch Of Spice	Casual Dining	https://www.zomato.com/agra/pinch-of-spice-taj...	1076/2, Fatehabad Road, Tajganj, Agra	Agra	34	Tajg
23	3401687	Jain Bhojnalaya	Bhojanalya	https://www.zomato.com/agra/jain-bhojnalaya-si...	Opposite ISBT, Transport Nagar, Sikandra, Agra	Agra	34	Sikan

	res_id	name	establishment	url	address	city	city_id	local
24	3400469	Papa Ji Da Dhaba	Dhaba	https://www.zomato.com/agra/papa-ji-da-dhaba-c...	Delhi Gate, Raja Mandi Station Road, Civil Lin...	Agra	34	Civil Li
25	3400021	Chokho Jeeman Marwari Jain Bhojanalya	Casual Dining	https://www.zomato.com/agra/chokho-jeeman-marw...	1/48, Delhi Gate, Station Road, Raja Mandi, Ci...	Agra	34	Civil Li
26	3400277	Pizza Hut	Casual Dining	https://www.zomato.com/agra/pizza-hut-1-civil-...	25, Deepshikha Towers, Sanjay Place, Civil Lin...	Agra	34	Civil Li
27	3400850	Urban Deck	Casual Dining	https://www.zomato.com/agra/urban-deck-2-civil...	5th Floor, The P L Palace Hotel, MG Road, Sanj...	Agra	34	Civil Li
28	3400251	Dominos Pizza	Quick Bites	https://www.zomato.com/agra/dominos-pizza-agra...	113-116, DC Nand Plaza, 50B Taj Road, Agra Can...	Agra	34	Agra Ca
29	18888066	Kiskey Whiskey	Bar	https://www.zomato.com/agra/kiskey-whiskey-3-c...	Floor 2, 119/8, Ashok Cosmos Mall, Sanjay Plac...	Agra	34	Civil Li
30	3400130	Manoj Restaurant	Quick Bites	https://www.zomato.com/agra/manoj-restaurant-m...	Kinari Bazaar, Hing Ki Mandi, Johri Bazaar, Ma...	Agra	34	Mant
31	3400475	Amritsari Kulcha Junction	Quick Bites	https://www.zomato.com/agra/amritsari-kulcha-j...	Deepshikha Building, Wazirpura tRoad, Sanjay P...	Agra	34	Civil Li
32	3400628	Deviram Food Circle	Sweet Shop	https://www.zomato.com/agra/deviram-food-circl...	Pushpanjali Commercial Complex, 23-24, Sikandr...	Agra	34	Lohama

	res_id	name	establishment	url	address	city	city_id	local
33	3400840	Subway	Quick Bites	https://www.zomato.com/agra/subway-khandari?ut...	Shop 37, Ground Floor, Omaxe SRK Mall, NH 2, N...	Agra	34	Khanc
34	3401344	Cafe Turquoise Cottage	Casual Dining	https://www.zomato.com/agra/cafe-turquoise-cot...	76, 2nd floor, Fatehabad Road, Near Amar Hotel...	Agra	34	Tajg
35	18895858	Guru Kripa Yadav Dhaba	Dhaba	https://www.zomato.com/agra/guru-kripa-yadav-d...	Delhi Gate, Raja Mandi Station Road, Civil Lin...	Agra	34	Civil Li
36	3400326	Tease Me - Rooftop Tea Boutique	Casual Dining	https://www.zomato.com/agra/tease-me-rooftop-t...	Near Purani Mandi Crossing,Fatehabad Road, Taj...	Agra	34	Tajg
37	3401492	Chimmanlal Puri Wale	Quick Bites	https://www.zomato.com/agra/chimmanlal-puri-wa...	Daresi Road, Jama Masjid, Kinari Bazar, Sheb B...	Agra	34	Mant
38	3401167	The Momo Corner	Quick Bites	https://www.zomato.com/agra/the-momo-corner-ka...	F-318 Kamla Nagar, Front of Sai baba Temple, n...	Agra	34	Kar Na
39	3400131	Jain Restaurant	Quick Bites	https://www.zomato.com/agra/jain-restaurant-sh...	110, Jaipur House Market, Opposite, Commercial...	Agra	34	Shahg

40 rows × 26 columns

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

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211944 entries, 0 to 211943
Data columns (total 26 columns):
 #   Column                Non-Null Count   Dtype  
---  ------                --------------   -----  
 0   res_id                211944 non-null  int64  
 1   name                  211944 non-null  object 
 2   establishment         207117 non-null  object 
 3   url                   211944 non-null  object 
 4   address               211810 non-null  object 
 5   city                  211944 non-null  object 
 6   city_id               211944 non-null  int64  
 7   locality              211944 non-null  object 
 8   latitude              211944 non-null  float64
 9   longitude             211944 non-null  float64
 10  zipcode               48757 non-null   object 
 11  country_id            211944 non-null  int64  
 12  locality_verbose      211944 non-null  object 
 13  cuisines              210553 non-null  object 
 14  timings               208070 non-null  object 
 15  average_cost_for_two  211944 non-null  int64  
 16  price_range           211944 non-null  int64  
 17  currency              211944 non-null  object 
 18  highlights            209875 non-null  object 
 19  aggregate_rating      211944 non-null  float64
 20  rating_text           211944 non-null  object 
 21  votes                 211944 non-null  int64  
 22  photo_count           211944 non-null  int64  
 23  opentable_support     211896 non-null  float64
 24  delivery              211944 non-null  int64  
 25  takeaway              211944 non-null  int64  
dtypes: float64(4), int64(9), object(13)
memory usage: 42.0+ MB
```

In [7]:
```
df.shape
```

Out[7]:
```
(211944, 26)
```

In [8]:
```
df.describe()
```

Out[8]:

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two
count	2.119440e+05	211944.000000	211944.000000	211944.000000	211944.0	211944.000000
mean	1.349411e+07	4746.785434	21.499475	77.615276	1.0	595.812229
std	7.883722e+06	5568.766386	22.781261	7.500104	0.0	606.239363
min	5.000000e+01	1.000000	0.000000	0.000000	1.0	0.000000
25%	3.301027e+06	11.000000	15.496071	74.877961	1.0	250.000000
50%	1.869573e+07	34.000000	22.514181	77.425971	1.0	400.000000
75%	1.881297e+07	11306.000000	26.841214	80.219323	1.0	700.000000
max	1.915979e+07	11354.000000	10000.000000	91.832769	1.0	30000.000000

In [10]:
```
df.columns = [c.strip().replace("","_") for c in df.columns]
```

In [11]:
```
df.isna()
```

Out[11]:

	__r__e__s_____i__d__	__n__a__m__e__	__e__s__t__a__b__l__i__s__h__m__e__n__t__
0	False	False	False
1	False	False	False
2	False	False	False
3	False	False	False
4	False	False	False
...	...	...	...
211939	False	False	False
211940	False	False	False
211941	False	False	False
211942	False	False	False
211943	False	False	False

211944 rows × 26 columns

In [21]:
```
df.head()
```

Out[21]:

	__r__e__s_____i__d__	__n__a__m__e__	__e__s__t__a__b__l__i__s__h__m__e__n__t__	
0	3400299	Bikanervala	Quick Bites	www agr
1	3400005	Mama Chicken Mama Franky House	Quick Bites	www chi
2	3401013	Bhagat Halwai	Quick Bites	www
3	3400290	Bhagat Halwai	Quick Bites	www
4	3401744	The Salt Cafe Kitchen & Bar	Casual Dining	www

5 rows × 26 columns

In [38]:
```
#correlation of data and save it in a variable
data.corr()
```

Out[38]:

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two	pr
res_id	1.000000	0.571824	0.019330	-0.041803	NaN	-0.208676	
city_id	0.571824	1.000000	0.031397	-0.038496	NaN	-0.273372	
latitude	0.019330	0.031397	1.000000	0.047109	NaN	-0.008643	
longitude	-0.041803	-0.038496	0.047109	1.000000	NaN	-0.019623	
country_id	NaN	NaN	NaN	NaN	NaN	NaN	
average_cost_for_two	-0.208676	-0.273372	-0.008643	-0.019623	NaN	1.000000	
price_range	-0.207115	-0.281038	-0.004729	-0.007658	NaN	0.785934	
aggregate_rating	-0.244654	-0.263028	-0.002171	0.069011	NaN	0.232347	
votes	-0.235851	-0.274275	-0.020725	-0.004032	NaN	0.269049	
photo_count	-0.106507	-0.236609	-0.013376	-0.021650	NaN	0.341905	
opentable_support	NaN	NaN	NaN	NaN	NaN	NaN	
delivery	0.008152	0.011540	0.011652	0.049892	NaN	-0.116311	
takeaway	NaN	NaN	NaN	NaN	NaN	NaN	

In [27]:
```
#plot the correlation
corr.style.background_gradient(cmap ='coolwarm')
```

```
C:\Users\gopal\anaconda3\lib\site-packages\pandas\io\formats\style.py:2813: RuntimeWa
rning: All-NaN slice encountered
  smin = np.nanmin(gmap) if vmin is None else vmin
C:\Users\gopal\anaconda3\lib\site-packages\pandas\io\formats\style.py:2814: RuntimeWa
rning: All-NaN slice encountered
  smax = np.nanmax(gmap) if vmax is None else vmax
```

Out[27]:

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two	pr
res_id	1.000000	0.571824	0.019330	-0.041803	nan	-0.208676	
city_id	0.571824	1.000000	0.031397	-0.038496	nan	-0.273372	
latitude	0.019330	0.031397	1.000000	0.047109	nan	-0.008643	
longitude	-0.041803	-0.038496	0.047109	1.000000	nan	-0.019623	
country_id	nan	nan	nan	nan	nan	nan	
average_cost_for_two	-0.208676	-0.273372	-0.008643	-0.019623	nan	1.000000	
price_range	-0.207115	-0.281038	-0.004729	-0.007658	nan	0.785934	
aggregate_rating	-0.244654	-0.263028	-0.002171	0.069011	nan	0.232347	

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two	pr
votes	-0.235851	-0.274275	-0.020725	-0.004032	nan	0.269049	
photo_count	-0.106507	-0.236609	-0.013376	-0.021650	nan	0.341905	
opentable_support	nan	nan	nan	nan	nan	nan	
delivery	0.008152	0.011540	0.011652	0.049892	nan	-0.116311	
takeaway	nan	nan	nan	nan	nan	nan	

In [35]:
```
df.isna()
```

Out[35]:

	res_id	name	establishment	url	address	city	city_id	locality	latitude	longitude	...
0	False	False	False	False	False	False	False	False	False	False	...
1	False	False	False	False	False	False	False	False	False	False	...
2	False	False	False	False	False	False	False	False	False	False	...
3	False	False	False	False	False	False	False	False	False	False	...
4	False	False	False	False	False	False	False	False	False	False	...
...	...	...	...	...	...	...	...	...	...	...	...
211939	False	False	False	False	False	False	False	False	False	False	...
211940	False	False	False	False	False	False	False	False	False	False	...
211941	False	False	False	False	False	False	False	False	False	False	...
211942	False	False	False	False	False	False	False	False	False	False	...
211943	False	False	False	False	False	False	False	False	False	False	...

211944 rows × 26 columns

In [41]:
```
print(df.isnull().sum())
```

```
res_id                   0
name                     0
establishment         4827
url                      0
address                134
city                     0
city_id                  0
locality                 0
latitude                 0
longitude                0
zipcode             163187
country_id               0
locality_verbose         0
cuisines              1391
timings               3874
average_cost_for_two     0
price_range              0
```

```
currency                    0
highlights               2069
aggregate_rating            0
rating_text                 0
votes                       0
photo_count                 0
opentable_support          48
delivery                    0
takeaway                    0
dtype: int64
```

In [43]:

```
updated_df=df.dropna(axis=1)
updated_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 211944 entries, 0 to 211943
Data columns (total 19 columns):
 #   Column                Non-Null Count   Dtype  
---  ------                --------------   -----  
 0   res_id                211944 non-null  int64  
 1   name                  211944 non-null  object 
 2   url                   211944 non-null  object 
 3   city                  211944 non-null  object 
 4   city_id               211944 non-null  int64  
 5   locality              211944 non-null  object 
 6   latitude              211944 non-null  float64
 7   longitude             211944 non-null  float64
 8   country_id            211944 non-null  int64  
 9   locality_verbose      211944 non-null  object 
 10  average_cost_for_two  211944 non-null  int64  
 11  price_range           211944 non-null  int64  
 12  currency              211944 non-null  object 
 13  aggregate_rating      211944 non-null  float64
 14  rating_text           211944 non-null  object 
 15  votes                 211944 non-null  int64  
 16  photo_count           211944 non-null  int64  
 17  delivery              211944 non-null  int64  
 18  takeaway              211944 non-null  int64  
dtypes: float64(3), int64(9), object(7)
memory usage: 30.7+ MB
```

In [44]:

```
updated_df1=df.dropna(axis=0)
updated_df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 47472 entries, 1 to 211943
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   res_id                47472 non-null  int64  
 1   name                  47472 non-null  object 
 2   establishment         47472 non-null  object 
 3   url                   47472 non-null  object 
 4   address               47472 non-null  object 
 5   city                  47472 non-null  object 
 6   city_id               47472 non-null  int64  
 7   locality              47472 non-null  object 
```

```
 8   latitude              47472 non-null  float64
 9   longitude             47472 non-null  float64
 10  zipcode               47472 non-null  object
 11  country_id            47472 non-null  int64
 12  locality_verbose      47472 non-null  object
 13  cuisines              47472 non-null  object
 14  timings               47472 non-null  object
 15  average_cost_for_two  47472 non-null  int64
 16  price_range           47472 non-null  int64
 17  currency              47472 non-null  object
 18  highlights            47472 non-null  object
 19  aggregate_rating      47472 non-null  float64
 20  rating_text           47472 non-null  object
 21  votes                 47472 non-null  int64
 22  photo_count           47472 non-null  int64
 23  opentable_support     47472 non-null  float64
 24  delivery              47472 non-null  int64
 25  takeaway              47472 non-null  int64
dtypes: float64(4), int64(9), object(13)
memory usage: 9.8+ MB
```

In [47]:
```
df.describe()
```

Out[47]:

	res_id	city_id	latitude	longitude	country_id	average_cost_for_two
count	2.119440e+05	211944.000000	211944.000000	211944.000000	211944.0	211944.000000
mean	1.349411e+07	4746.785434	21.499475	77.615276	1.0	595.812229
std	7.883722e+06	5568.766386	22.781261	7.500104	0.0	606.239363
min	5.000000e+01	1.000000	0.000000	0.000000	1.0	0.000000
25%	3.301027e+06	11.000000	15.496071	74.877961	1.0	250.000000
50%	1.869573e+07	34.000000	22.514181	77.425971	1.0	400.000000
75%	1.881297e+07	11306.000000	26.841214	80.219323	1.0	700.000000
max	1.915979e+07	11354.000000	10000.000000	91.832769	1.0	30000.000000

In [72]:

```
# create region variable/ column/ feature

if "latitude" in df.columns and "longitude" in df.columns:
    lat_med=df["latitude"].median()
    lon_med=df["longitude"].median()
def get_region(lat,lon):
    if pd.isna(lat) or pd.isna(lon):
        return"unknown"
    if lat>=lat_med and lon>=lon_med:
        return "NE"
    if lat>=lat_med and lon<=lon_med:
        return"NW"
    if lat<=lat_med and lon>=lon_med:
        return"SE"
    else:
        return"SW"

df["region"]=df.apply(lambda row: get_region(row["latitude"],row["longitude"]),axis=1
df.head(211944)
```

Out[72]:

	res_id	name	establishment	url	address	city	city_id	loca
0	3400299	Bikanervala	Quick Bites	https://www.zomato.com/agra/bikanervala-khanda...	Kalyani Point, Near Tulsi Cinema, Bypass Road,...	Agra	34	Khan
1	3400005	Mama Chicken Mama Franky House	Quick Bites	https://www.zomato.com/agra/mama-chicken-mama-...	Main Market, Sadar Bazaar, Agra Cantt, Agra	Agra	34	A C
2	3401013	Bhagat Halwai	Quick Bites	https://www.zomato.com/agra/bhagat-halwai-2-sh...	62/1, Near Easy Day, West Shivaji Nagar, Goalp...	Agra	34	Shahg
3	3400290	Bhagat Halwai	Quick Bites	https://www.zomato.com/agra/bhagat-halwai-civi...	Near Anjana Cinema, Nehru Nagar, Civil Lines, ...	Agra	34	Civil L

	res_id	name	establishment	url	address	city	city_id	loca
4	3401744	The Salt Cafe Kitchen & Bar	Casual Dining	https://www.zomato.com/agra/the-salt-cafe-kitc...	1C,3rd Floor, Fatehabad Road, Tajganj, Agra	Agra	34	Tajg
...	...	...	...	...	...	...	...	
211939	3202251	Kali Mirch Cafe And Restaurant	Casual Dining	https://www.zomato.com/vadodara/kali-mirch-caf...	Manu Smriti Complex, Near Navrachna School, Gl...	Vadodara	32	Fatehg
211940	3200996	Raju Omlet	Quick Bites	https://www.zomato.com/vadodara/raju-omlet-kar...	Mahalaxmi Apartment, Opposite B O B, Karoli Ba...	Vadodara	32	Karelib
211941	18984164	The Grand Thakar	Casual Dining	https://www.zomato.com/vadodara/the-grand-thak...	3rd Floor, Shreem Shalini Mall, Opposite Conqu...	Vadodara	32	Alka
211942	3201138	Subway	Quick Bites	https://www.zomato.com/vadodara/subway-1-akota...	G-2, Vedant Platina, Near Cosmos, Akota, Vadodara	Vadodara	32	Ak
211943	18879846	Freshcos - The Health Cafe	Café	https://www.zomato.com/vadodara/freshcos-the-h...	Shop 7, Ground Floor, Opposite Natubhai Circle...	Vadodara	32	Vadiv

211944 rows × 27 columns

In [74]:
```
if "latitude" in df.columns and "longitude" in df.columns:
    lat_med=df["latitude"].median()
    lon_med=df["longitude"].median()
def get_region(lat,lon):
    if pd.isna(lat) or pd.isna(lon):
        return"unknown"
    if lat>=lat_med and lon>=lon_med:
        return "NE"
    if lat>=lat_med and lon<=lon_med:
        return"NW"
    if lat<=lat_med and lon>=lon_med:
        return"SE"
    else:
        return"SW"

df["region"]=df.apply(lambda row: get_region(row["latitude"],row["longitude"]),axis=1

df["region"].value_counts()
```

Out[74]:
```
NE    54990
SW    54980
NW    50992
SE    50982
Name: region, dtype: int64
```

In [76]:
```
import numpy as np

# detect outliers using IQR
def detect_outliers(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    lower_bound = Q1 - 1.5 * IQR
    upper_bound = Q3 + 1.5 * IQR
    outliers = df[(df[column] < lower_bound) | (df[column] > upper_bound)][column]
    return outliers
#key numeric columns ( latitude, longitude, average_cost_for_two, aggregate_rating)
columns_to_check = ['latitude', 'longitude', 'average_cost_for_two', 'aggregate_ratir

for column in columns_to_check:
    outliers = detect_outliers(df, column)
    print(f"Outliers in {column}:")
    print(outliers)
    print(f"Number of outliers: {len(outliers)}\n")


plt.figure(figsize=(12, 6))
sns.boxplot(data=df[columns_to_check])
plt.title('Boxplot of Numeric Variables')
plt.xticks(rotation=45)
plt.show()
```

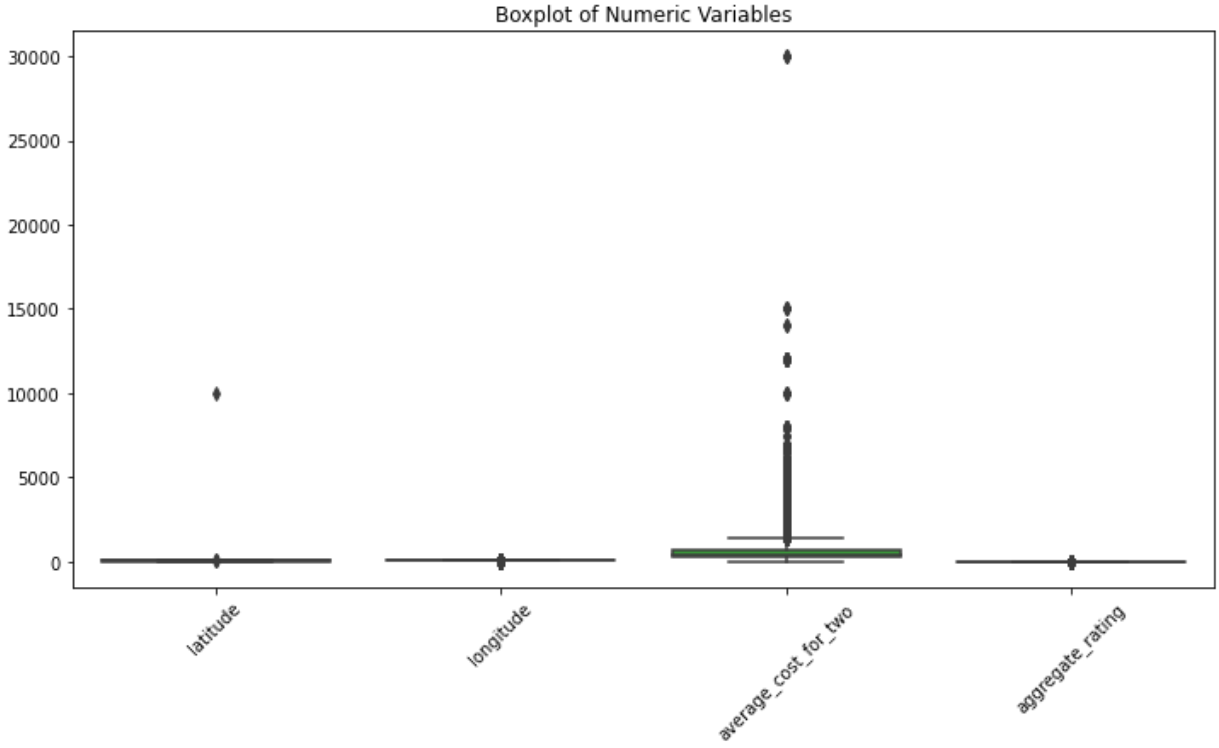
```
Outliers in latitude:
86942        75.791315
88299        75.791315
88596        75.791315
125880    10000.000000
```

```
Name: latitude, dtype: float64
Number of outliers: 4

Outliers in longitude:
111       0.0
113       0.0
122       0.0
145       0.0
149       0.0
         ...
209750    0.0
209755    0.0
209760    0.0
209762    0.0
209764    0.0
Name: longitude, Length: 11487, dtype: float64
Number of outliers: 11487

Outliers in average_cost_for_two:
16        1600
22        1600
27        1600
34        1600
42        1500
          ...
211875    2000
211890    1400
211892    1500
211893    2000
211894    1400
Name: average_cost_for_two, Length: 20017, dtype: int64
Number of outliers: 20017

Outliers in aggregate_rating:
103       0.0
114       0.0
132       0.0
133       0.0
134       0.0
         ...
210590    0.0
210592    0.0
210593    0.0
210594    0.0
210802    0.0
Name: aggregate_rating, Length: 23518, dtype: float64
Number of outliers: 23518
```

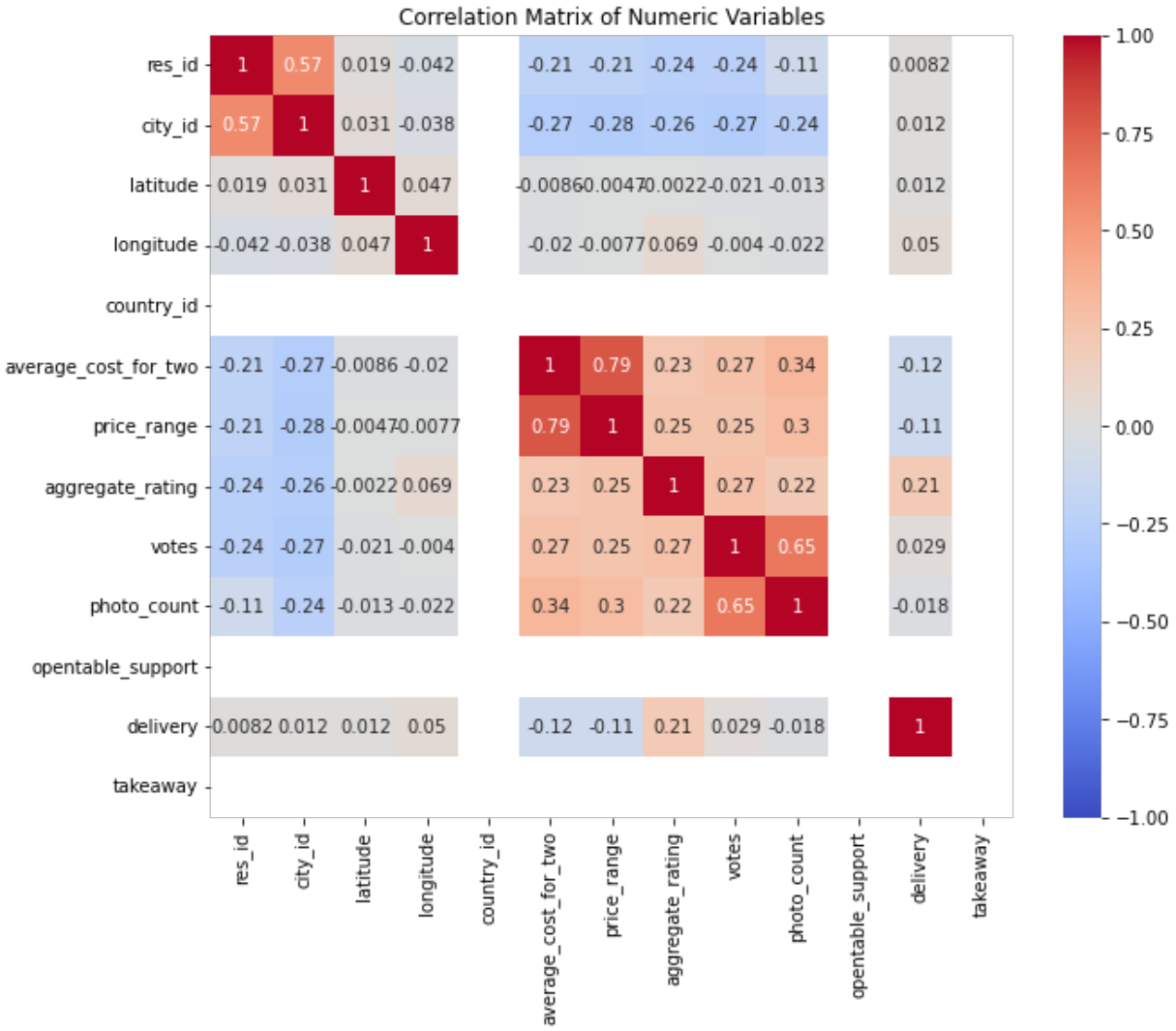


In [78]:

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Select only numeric columns
numeric_df = df.select_dtypes(include=['float64', 'int64'])

# Calculate correlation matrix
correlation_matrix = numeric_df.corr()
#correlation matrix with a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', vmin=-1, vmax=1)
plt.title('Correlation Matrix of Numeric Variables')
plt.show()
```



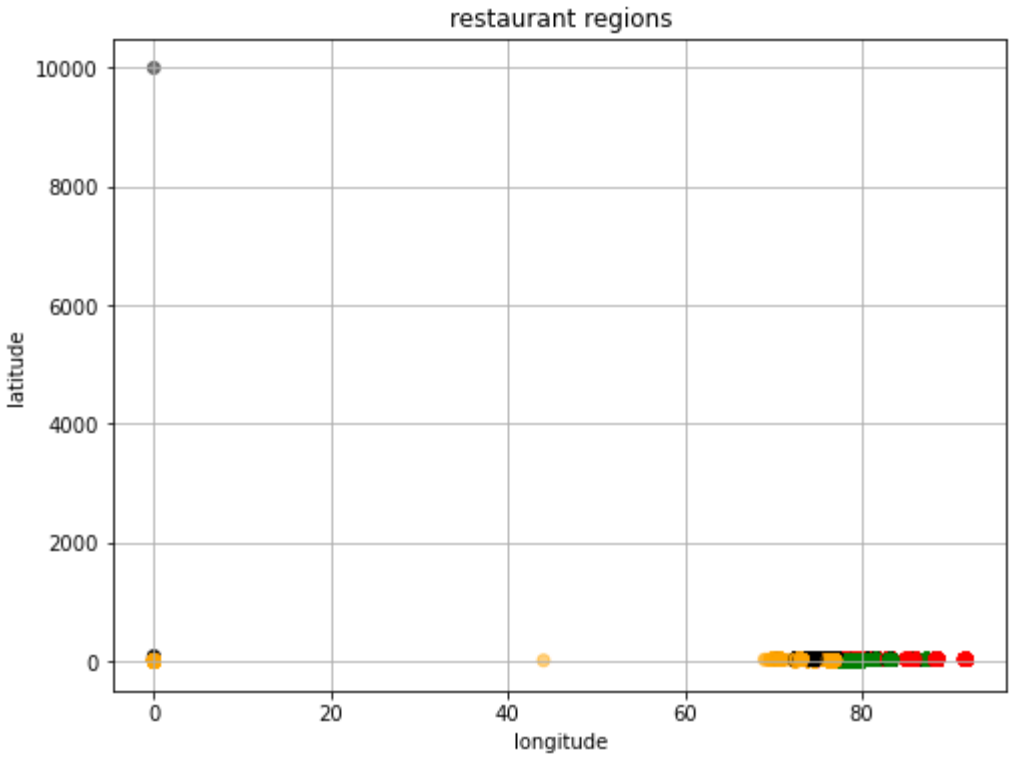
In [82]:

```
import matplotlib.pyplot as plt
import seaborn as sns

color_map ={'NE':'red','NW':'black','SE':'green','SW':'orange','unkown':'gray}'}
colors = df['region'].map(color_map)

plt.figure(figsize=(8,6))
plt.scatter(df['longitude'],df['latitude'],c=colors,alpha=0.5)

plt.title('restaurant regions')
plt.xlabel('longitude')
plt.ylabel('latitude')
plt.grid(True)
plt.show()
```



In [84]:

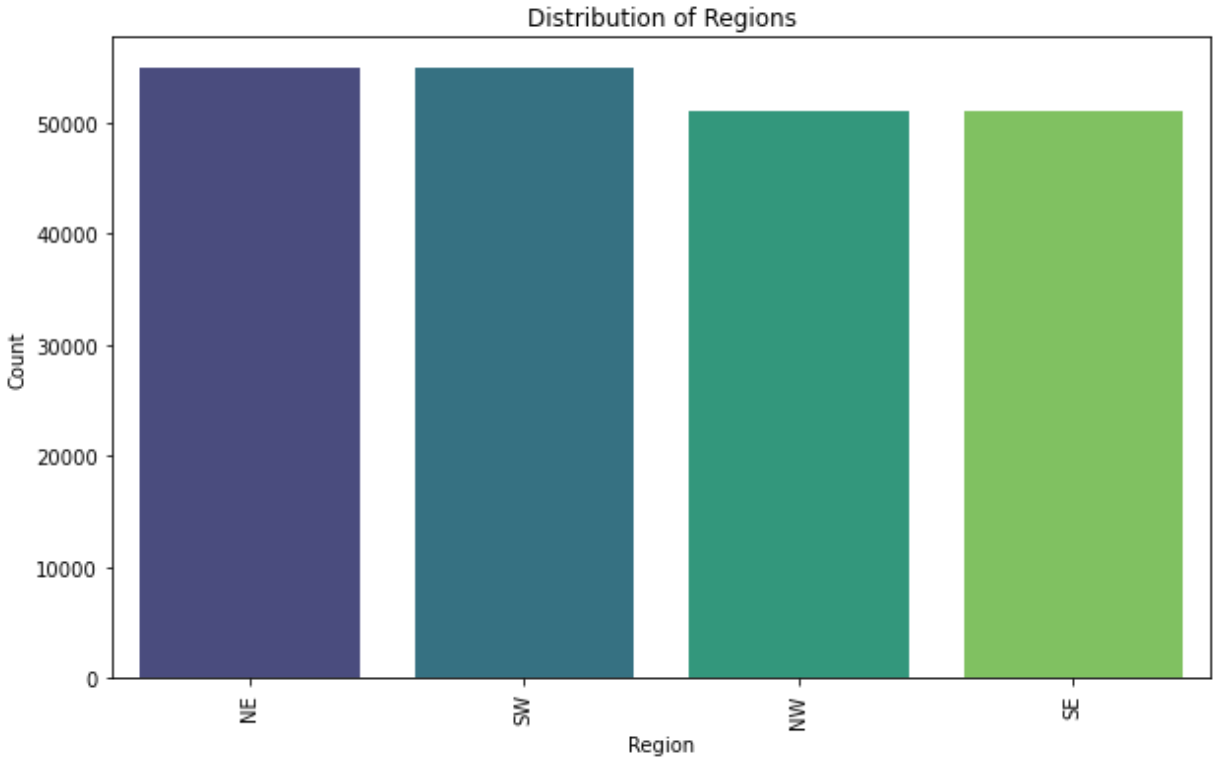
```
import matplotlib.pyplot as plt
import seaborn as sns

region_counts = df['region'].value_counts()
plt.figure(figsize=(10, 6))

sns.barplot(x=region_counts.index, y=region_counts.values, palette='viridis')


plt.title('Distribution of Regions')
plt.xlabel('Region')
plt.ylabel('Count')
plt.xticks(rotation=90)

plt.show()
```


Distribution of Regions
Count
50000
40000
30000
20000
10000
0
NE
SW
NW
SE
Region

In [ ]: