

Group10_Visualizations

January 21, 2021

```
[1]: import matplotlib.pyplot as plt
import numpy as np
import nltk
from wordcloud import WordCloud
```

0.1 Vaishnavi - Top 5 Cities with Most Yelp Reviews

```
[2]: cities = ['Las Vegas', 'Phoenix', 'Toronto', 'Scottsdale', 'Charlotte']
reviews = [2360735, 842321, 583512, 439439, 371580]
```

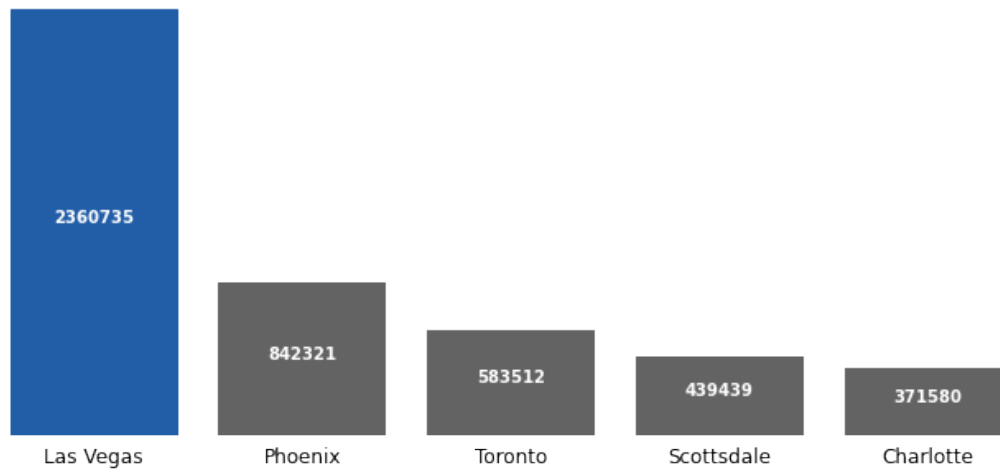
```
[3]: fig, ax = plt.subplots(figsize=(12, 5))
barcontainers = ax.bar(cities, reviews, color="#636363")
barcontainers[0].set_color('#225ea8')

ax.set_title("Top 5 Cities with Most Yelp Reviews",
             loc='left',
             fontweight='bold',
             fontsize=16,
             pad=50)
ax.get_yaxis().set_visible(False)
ax.xaxis.set_ticks_position('none')
ax.tick_params(labelsize=12)
ax.spines['top'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['bottom'].set_visible(False)
ax.spines['left'].set_visible(False)

for p in ax.patches:
    width = p.get_width()
    height = p.get_height()
    x, y = p.get_xy()
    ax.annotate(f'{height}', (x + width / 2, y + height * 0.5),
               ha='center',
               fontweight='bold',
               color='white')

plt.show()
```

Top 5 Cities with Most Yelp Reviews



0.2 April - Average Review Stars by US State

```
[4]: plot_data = {  
    'CT': '4.50',  
    'VT': '4.00',  
    'CA': '3.85',  
    'AL': '3.83',  
    'NV': '3.65',  
    'AZ': '3.65',  
    'WI': '3.57',  
    'PA': '3.55',  
    'AR': '3.50',  
    'MI': '3.50'  
}  
labels = list(plot_data.keys())  
values = [float(x) for x in plot_data.values()]  
  
[5]: fig, ax = plt.subplots(figsize=(6, 8))  
y_pos = np.arange(len(labels))  
  
ax.barh(y_pos, values, color="#01665e", height=0.66)  
ax.invert_yaxis()  
ax.set_xticks([])  
ax.set_yticks(y_pos)  
ax.set_yticklabels(labels, fontsize=13, fontweight="bold")  
ax.set_title("Average Review Stars Over States",  
            fontsize=16,
```

```

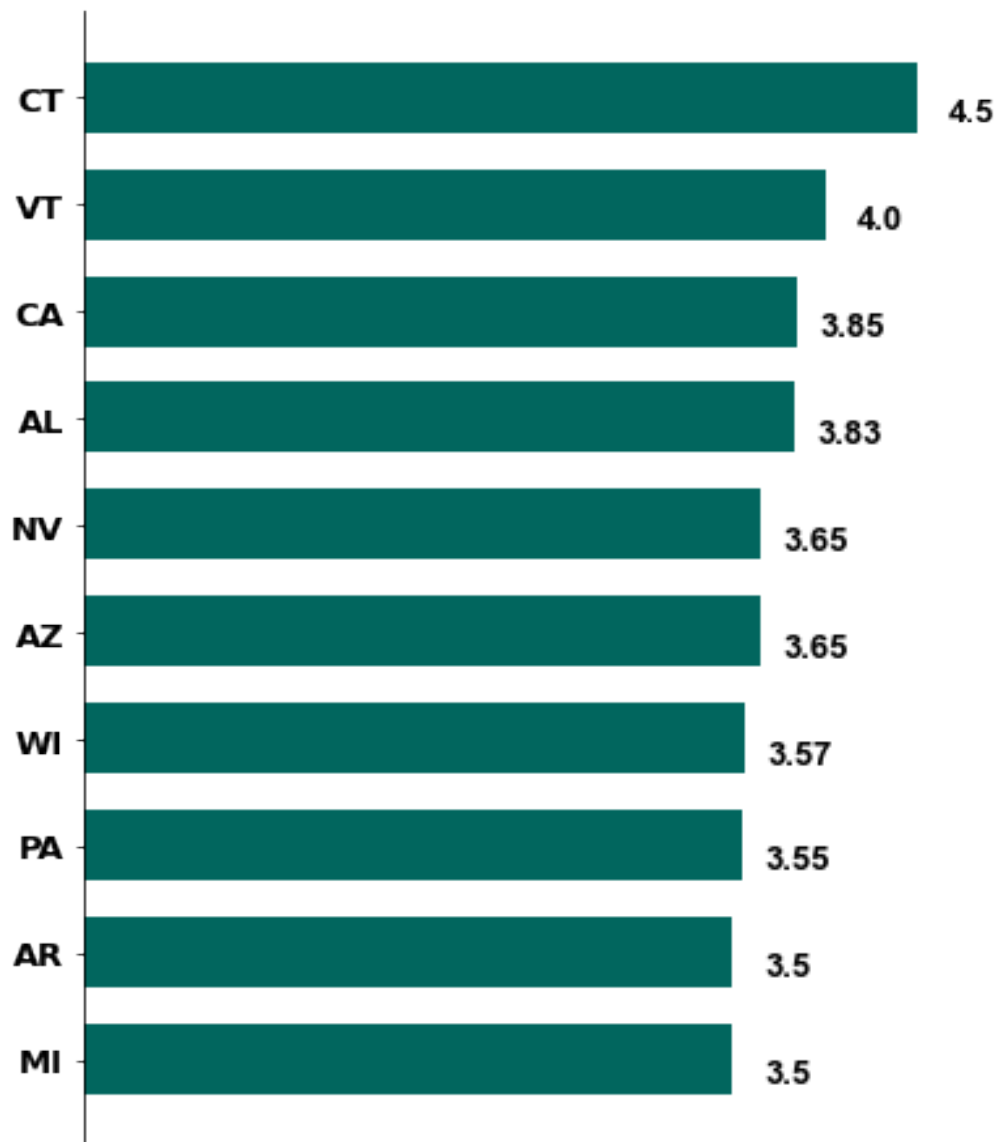
        fontname="Arial",
        fontweight="bold",
        ha="left",
        va="top",
        x=-0.2,
        y=1.1)
ax.spines['bottom'].set_visible(False)
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)

for i, v in enumerate(values):
    ax.text(v + 0.3,
            i,
            v,
            fontname="Arial",
            ha="center",
            va="top",
            fontweight="bold",
            fontsize=13)

plt.show()

```

Average Review Stars Over States



0.3 April - Word Cloud Plot

```
[6]: data_emr = [('food', 1052690), ('just', 956295), ('service', 893304),  
               ('place', 878106), ('time', 871369), ('like', 845309),  
               ('good', 663843), ('said', 616255), ('don', 586809),  
               ('told', 571192), ('got', 539984), ('did', 508988),  
               ('order', 500890), ('asked', 459766), ('came', 458860),  
               ('went', 444762), ('ordered', 425525), ('minutes', 423529),
```

```
(
    'really', 421045), ('people', 408384), ('going', 357492),
    ('know', 356631), ('customer', 345264), ('come', 336684),
    ('bad', 323107), ('took', 322067), ('experience', 315648),
    ('way', 314923), ('called', 312193), ('restaurant', 302192),
    ('make', 301644), ('day', 299967), ('better', 299828),
    ('manager', 291620), ('great', 286047), ('want', 283980),
    ('car', 273963), ('staff', 271359), ('wait', 264551),
    ('say', 263577), ('left', 261655), ('money', 255375),
    ('room', 254878), ('wasn', 253968), ('table', 250987),
    ('new', 249925), ('wanted', 242952), ('chicken', 242019),
    ('times', 239779), ('work', 236489), ('nice', 228080),
    ('right', 228042), ('location', 227344), ('think', 213069),
    ('rude', 211656), ('worst', 211113), ('business', 206807),
    ('long', 206016), ('try', 205553), ('pay', 203424),
    ('night', 200337), ('sure', 199630), ('check', 194976),
    ('price', 194501), ('away', 193183), ('store', 192027),
    ('hour', 188383), ('eat', 184685), ('ask', 182566),
    ('waiting', 181677), ('phone', 180679), ('server', 180451),
    ('need', 180272), ('horrible', 179879), ('home', 179774),
    ('little', 178127), ('company', 177914), ('later', 175525),
    ('pretty', 172816), ('bar', 172446), ('finally', 171850),
    ('menu', 170773), ('let', 170460), ('pizza', 167298),
    ('tried', 166699), ('tell', 166339), ('thing', 164013),
    ('care', 158265), ('gave', 156251), ('terrible', 155971),
    ('waited', 154533), ('disappointed', 154450), ('look', 154221),
    ('customers', 153769), ('looked', 153116), ('person', 151317),
    ('hours', 149565), ('drinks', 148827), ('small', 148535)]
```

```
data_emr = dict(data_emr)
words = data_emr.keys()
counts = data_emr.values()
pos = dict(nltk.pos_tag(list(data_emr.keys())))
nouns = {}
for word in words:
    if (pos[word] == "NN") and (pos[word] != "wasn"):
        nouns[word] = data_emr[word]
wordcloud = WordCloud()
wordcloud.fit_words(nouns)
fig = plt.figure(figsize=(10, 20))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



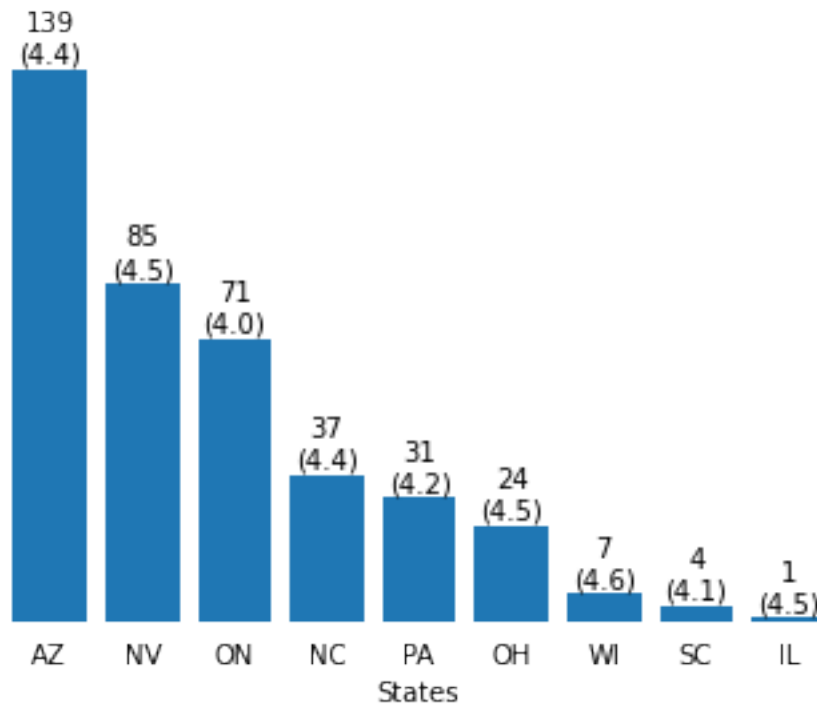
```

ax.text(x=-1,
        y=-35,
        s="Note: Figures in parenthesis represent average ratings")
ax.set_xlabel("States")
ax.set_ylabel("Count")
ax.set_title("Pilates Studio Count and Average Rating by State in the US",
             pad=25)
ax.get_yaxis().set_visible(False)
plt.tick_params(axis='x',
                which='both',
                bottom=False,
                top=False,
                labelbottom=True)

for spine in ax.spines.values():
    spine.set_visible(False)
plt.show()

```

Pilates Studio Count and Average Rating by State in the US



Note: Figures in parenthesis represent average ratings

0.5 Elyse - Yelp Reviews Over the Years

```
[9]: # Total Reviews, Avg Rating, Total # Users
data = [('2004', (12369, 3.86, 78)), ('2005', (244748, 3.82, 1016)),
        ('2006', (1012698, 3.76, 5956)), ('2007', (2207288, 3.75, 16737)),
        ('2008', (3390969, 3.74, 33116)), ('2009', (4471667, 3.74, 64981)),
        ('2010', (5412524, 3.75, 108955)), ('2011', (5984456, 3.75, 172808)),
        ('2012', (4828564, 3.71, 186580)), ('2013', (4169659, 3.69, 201430)),
        ('2014', (3694694, 3.68, 227023)), ('2015', (2926666, 3.66, 240722)),
        ('2016', (1929175, 3.63, 215620)), ('2017', (1018181, 3.58, 140627)),
        ('2018', (553787, 3.49, 107100)), ('2019', (202814, 3.27, 64251))]

d = {}
for a, b in data:
    d[int(a)] = b
reviews = [x[0] for x in d.values()]
rating = [x[1] for x in d.values()]
users = [x[2] for x in d.values()]
```

```
[10]: fig, ax = plt.subplots(figsize=(8, 6))
ax.get_yaxis().get_major_formatter().set_scientific(False)
rects = ax.bar(d.keys(), reviews, color='#0E3C50')
fig.suptitle('Yelp Reviews through the Years', fontsize=16, x=.2, y=1)
plt.xlim([2003, 2019.5])
plt.yticks(np.arange(0, 6000001, 1000000), color='#2D3234')
plt.ylim([0, 6200000])
ax.set_yticklabels([0, 1, 2, 3, 4, 5, 6])

def autolabel(rects):
    """Attach a text label above each bar in *rects*, displaying its height."""
    for i, rect in enumerate(rects):
        height = rect.get_height()
        ax.annotate(
            '{}'.format(round(users[i] / 1000)),
            xy=(rect.get_x() + rect.get_width() / 2, height),
            xytext=(0, 3), # 3 points vertical offset
            textcoords="offset points",
            ha='center',
            va='bottom',
            color='#383B40')

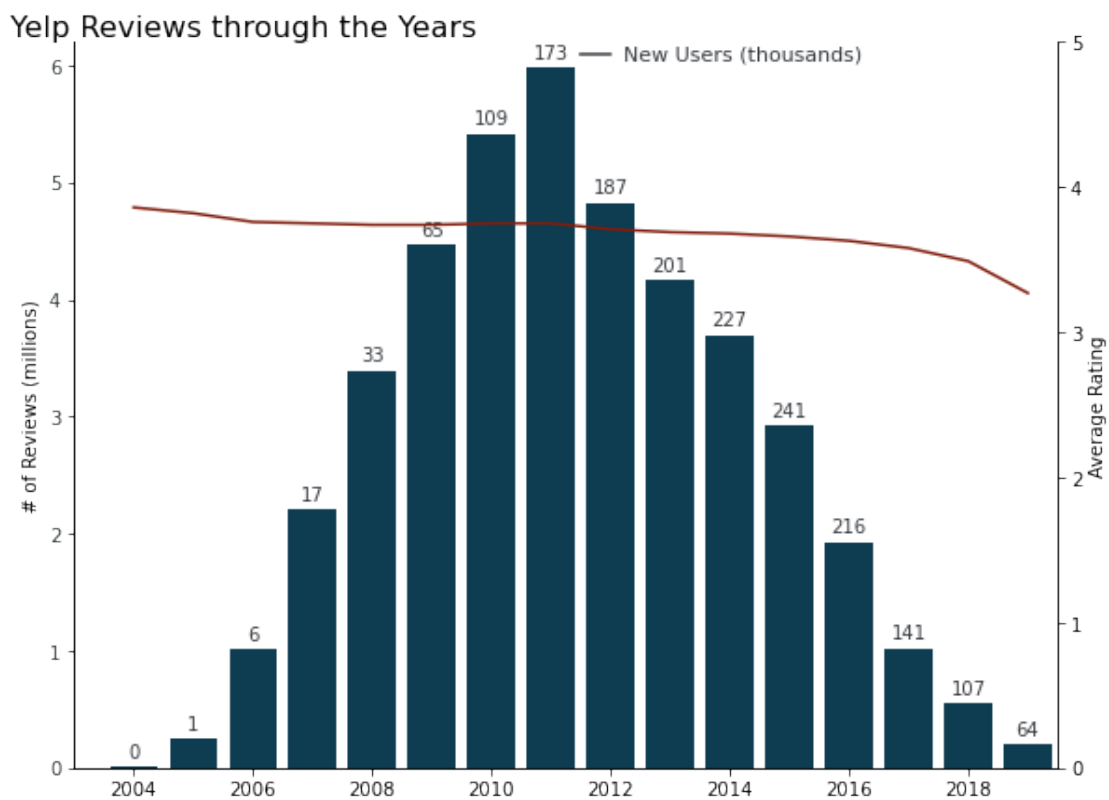
plt.text(2012.2,
        6050000,
        'New Users (thousands)',
        fontsize=11,
        color='#383B40')
```



```

plt.plot([2011.5, 2012], [6100000, 6100000], c='#383B40')
autolabel(rects)
ax2 = ax.twinx()
ax2.plot(list(d.keys()), rating, color='#821E0F')
ax2.set_ylim(0, 5)
fig.tight_layout()
ax.set_ylabel('# of Reviews (millions)')
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax2.spines['top'].set_visible(False)
ax2.set_ylabel('Average Rating')
plt.show()

```



0.6 Sophie - Categories with Top Review Counts in CA

```

[11]: category = [
        'Pizza, Italian', 'Arts & Entertainment',
        'Automotive, Roadside Assistance', 'Event Planning & Services',
        'Burgers, Fast Food'
    ]
review_count = [59, 35, 34, 34, 28]

```

```
colors = ["#d68db3"] * 5
colors[0] = "#d5a4eb"
```

```
[12]: fig, ax = plt.subplots(figsize=(10, 5))
ax.bar(category, review_count, width=0.5, color=colors, edgecolor='black')

ax.set_ylabel("Review Count", fontweight='bold')
plt.xticks(rotation=45)
ax.spines['right'].set_visible(False)
ax.spines['top'].set_visible(False)
ax.set_title("Business Categories with Top Review Counts in CA",
             fontweight='bold',
             size=18)

for i, v in enumerate(review_count):
    ax.text(i - 0.11,
            v / review_count[i],
            review_count[i],
            fontsize=18,
            color='black')
```

