Arizona City Weather Data

By: Jake Klingler

Abstract

The goal of this project was to use selenium to crawl over city data urls, find images within the city data pages, download the images, extract information from images, and store the tables in a SQL database. Then to create visualizations to better understand and analyze the data.

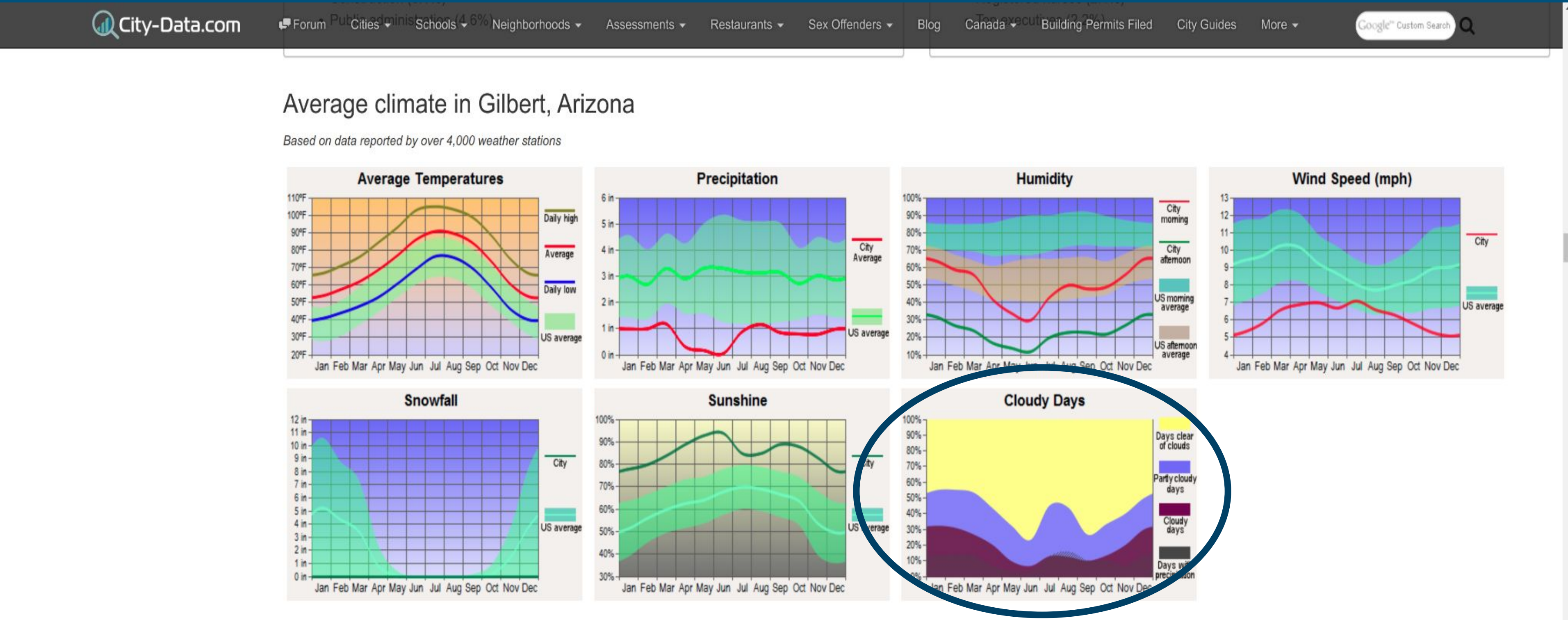
Background

- Flagstaff, Gilbert, and Tucson are the 3 AZ cities that were used
- Website that was used was city-data.com
    - Has data for every city in US
- The weather data contains:
    - Sunny days
    - Partly cloudy
    - Fully cloudy
    - Precipitation

Analysis Approach

- Crawl over city data urls via Selenium
- Find weather images in page source
- Download the found images
- Programmatically retrieve data from image
- Visualize the data
- Compare and Analyze the data for each city

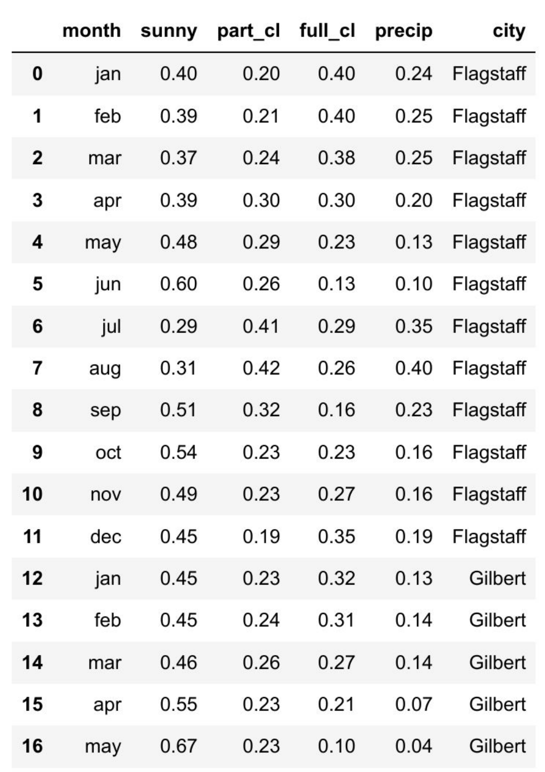
City Data Website



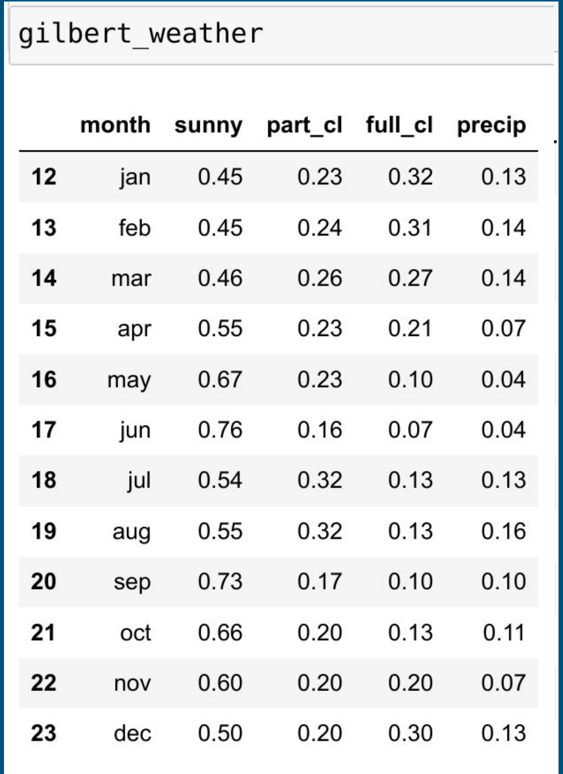
Average climate in Gilbert, Arizona

Based on data reported by over 4,000 weather stations

Cleaning Data



	month	sunny	part_cl	full_cl	precip	city
0	jan	0.40	0.20	0.40	0.24	Flagstaff
1	feb	0.39	0.21	0.40	0.25	Flagstaff
2	mar	0.37	0.24	0.38	0.25	Flagstaff
3	apr	0.39	0.30	0.30	0.20	Flagstaff
4	may	0.48	0.29	0.23	0.13	Flagstaff
5	jun	0.60	0.26	0.13	0.10	Flagstaff
6	jul	0.29	0.41	0.29	0.35	Flagstaff
7	aug	0.31	0.42	0.26	0.40	Flagstaff
8	sep	0.51	0.32	0.16	0.23	Flagstaff
9	oct	0.54	0.23	0.23	0.16	Flagstaff
10	nov	0.49	0.23	0.27	0.16	Flagstaff
11	dec	0.45	0.19	0.35	0.19	Flagstaff
12	jan	0.45	0.23	0.32	0.13	Gilbert
13	feb	0.45	0.24	0.31	0.14	Gilbert
14	mar	0.46	0.26	0.27	0.14	Gilbert
15	apr	0.55	0.23	0.21	0.07	Gilbert
16	may	0.67	0.23	0.10	0.04	Gilbert

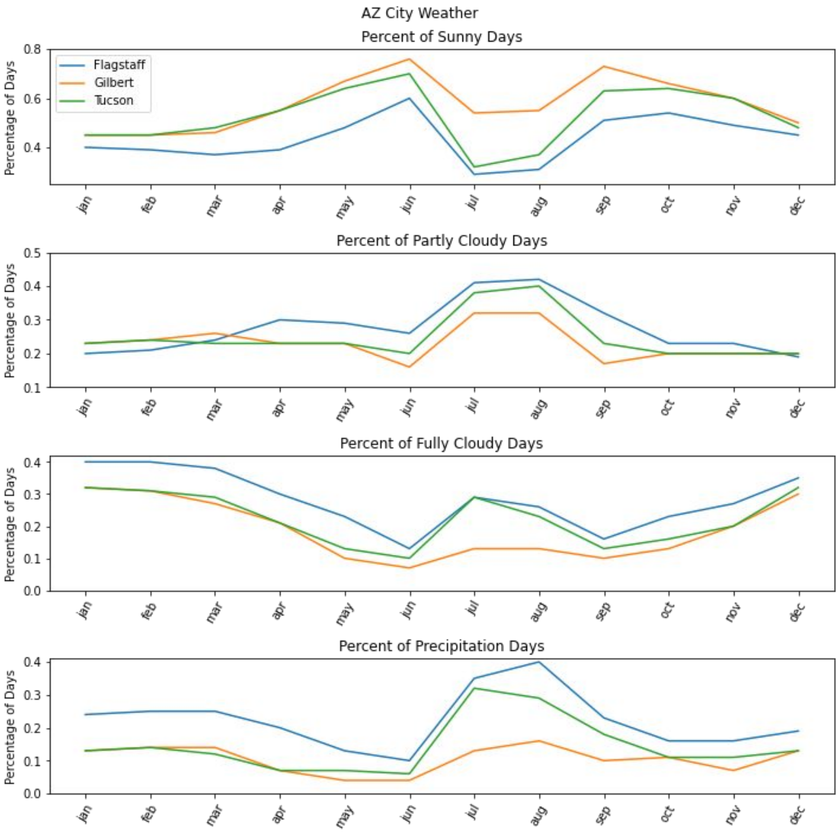


```
gilbert_weather
```

	month	sunny	part_cl	full_cl	precip
12	jan	0.45	0.23	0.32	0.13
13	feb	0.45	0.24	0.31	0.14
14	mar	0.46	0.26	0.27	0.14
15	apr	0.55	0.23	0.21	0.07
16	may	0.67	0.23	0.10	0.04
17	jun	0.76	0.16	0.07	0.04
18	jul	0.54	0.32	0.13	0.13
19	aug	0.55	0.32	0.13	0.16
20	sep	0.73	0.17	0.10	0.10
21	oct	0.66	0.20	0.13	0.11
22	nov	0.60	0.20	0.20	0.07
23	dec	0.50	0.20	0.30	0.13

Exploratory Data Analysis

- See similar shapes
- Monsoon season
  - June 15th- Sep. 30th
- Tucson had largest % change
  - Sunny days decrease by 38% June-July
  - Precipitation increase by 26% June-July



AZ City Weather

Percent of Sunny Days

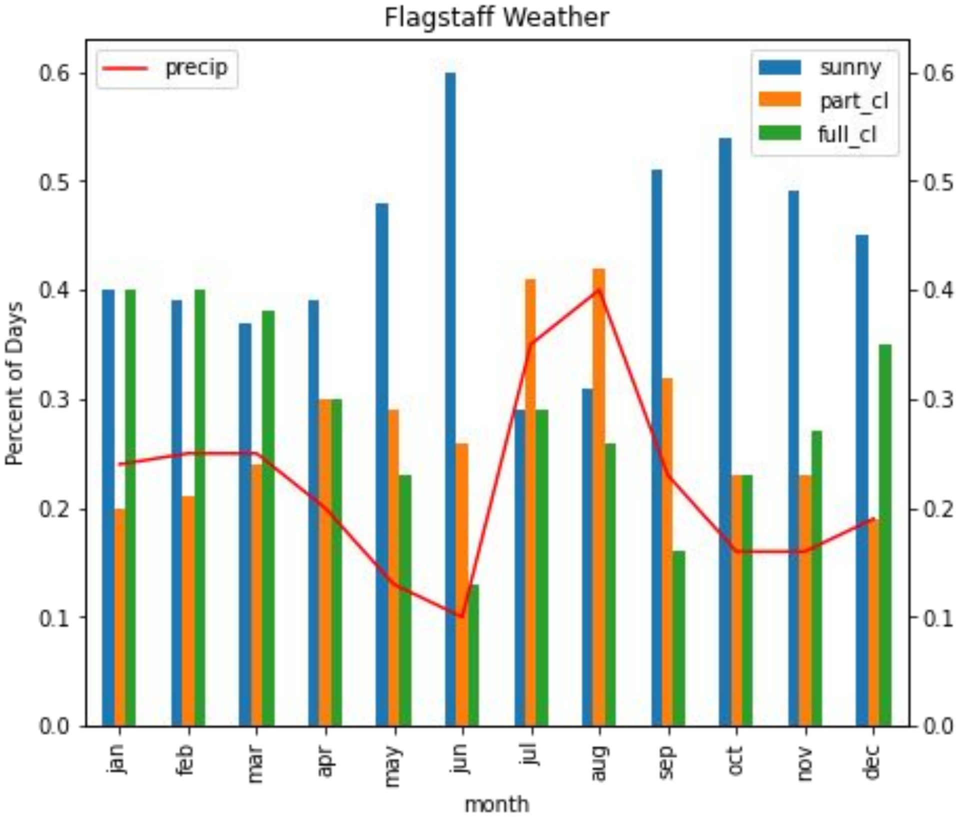
Percent of Partly Cloudy Days

Percent of Fully Cloudy Days

Percent of Precipitation Days

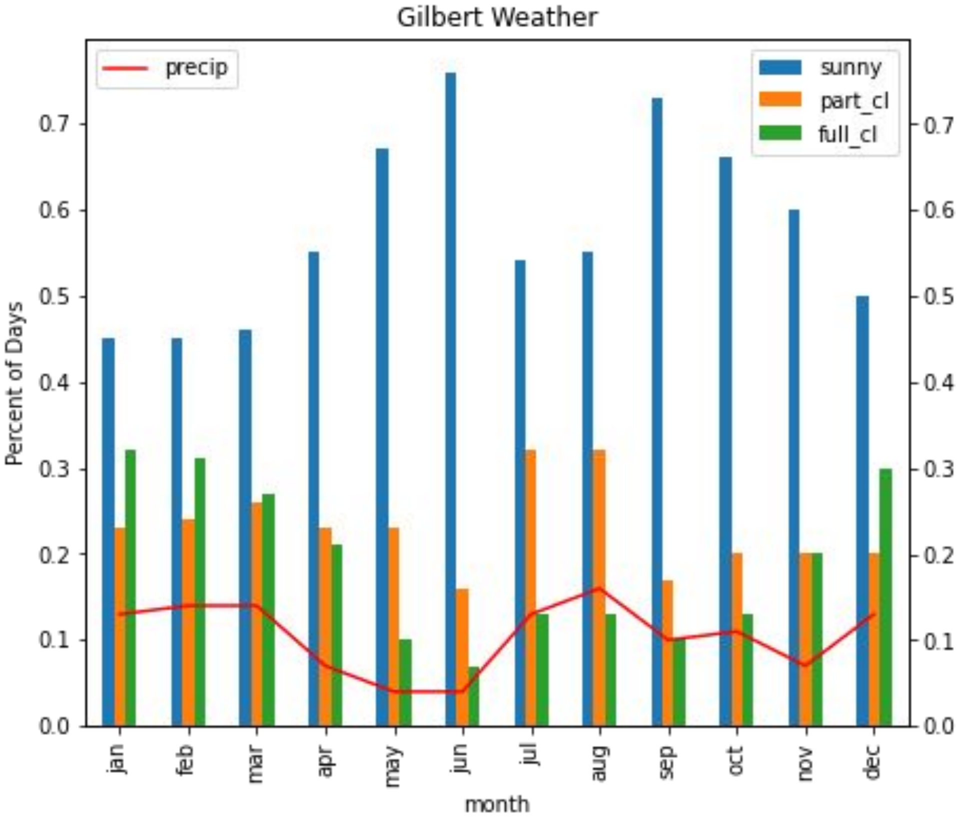
Analysis Cont.

- See which weather category dominates each month
- Monsoon season still noticeable
- Precipitation is overlaid as line plot
- Sunny+part_cl+full_cl = 100%



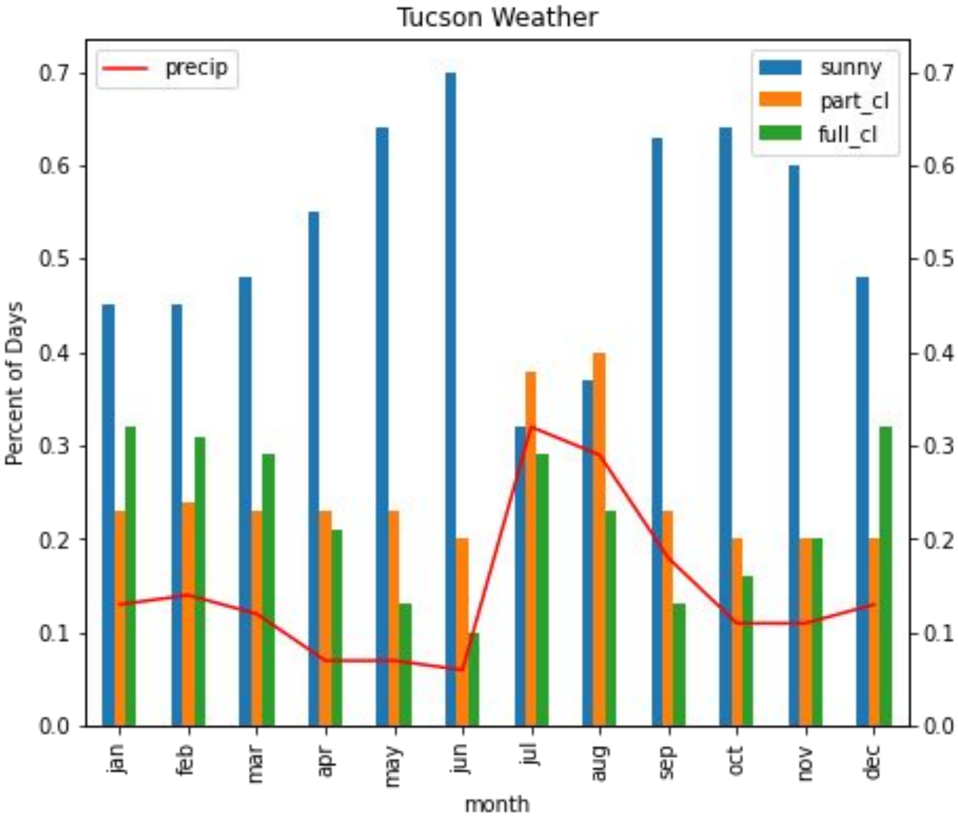
Analysis Cont.

- We see that all 12 months are dominated by sunny days by pretty large margin.
- 76% of days in June were sunny (largest % of any month for any city)

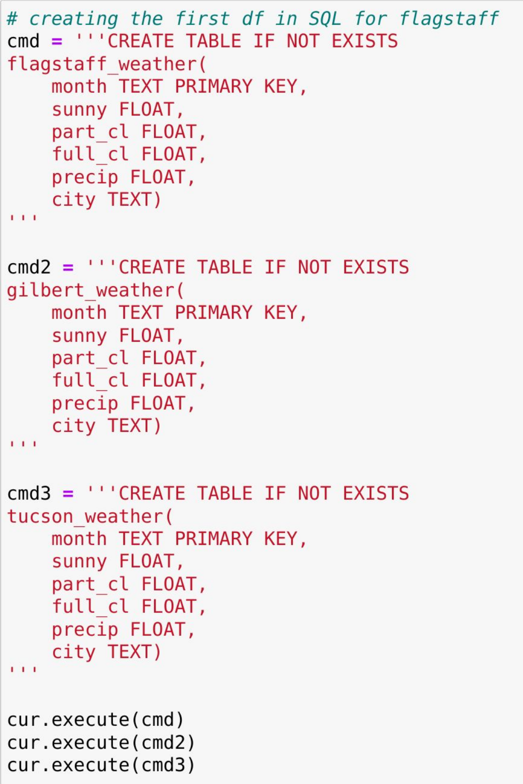


Analysis Cont.

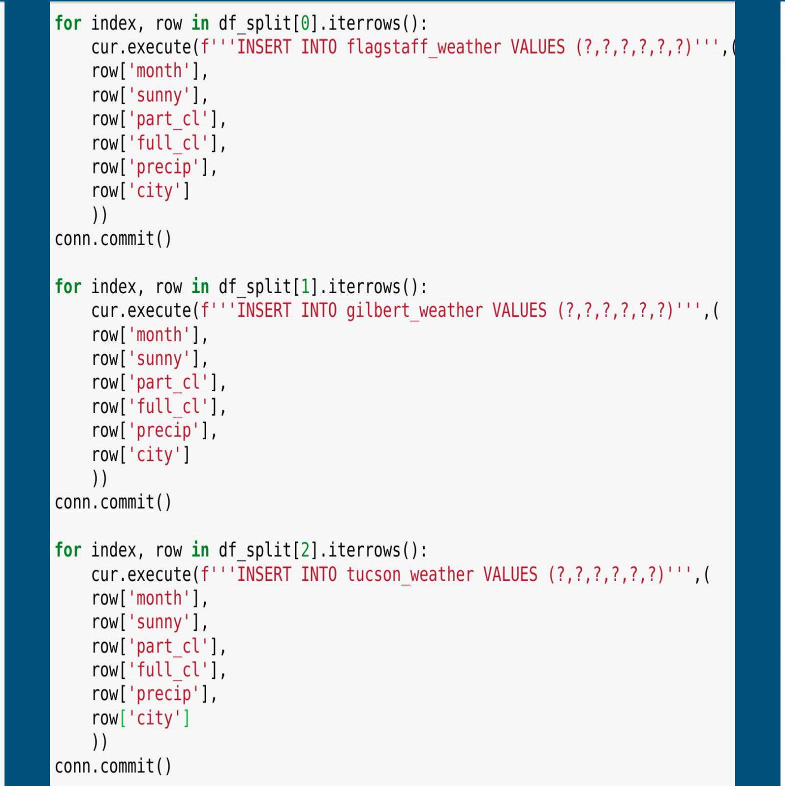
- 10/12 months are dominated by sunny days



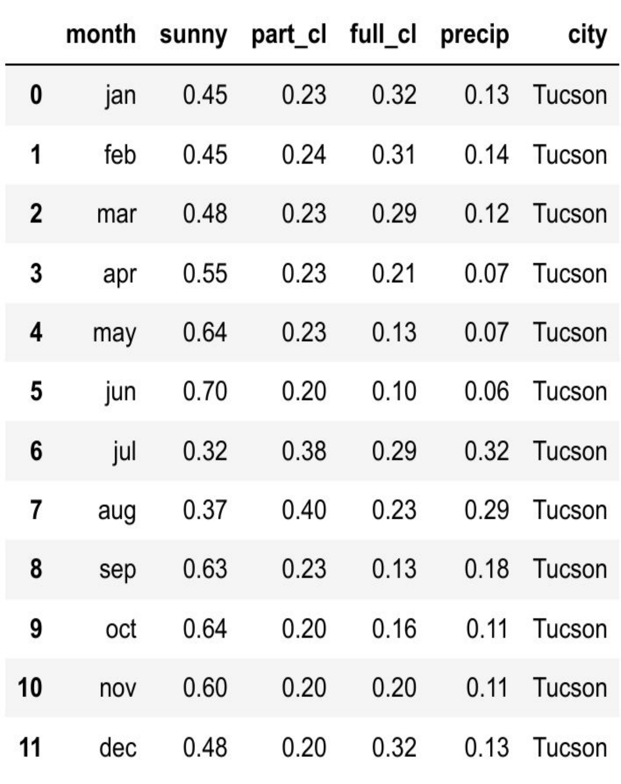
SQL DataBase



```
# creating the first df in SQL for flagstaff
cmd = '''CREATE TABLE IF NOT EXISTS
flagstaff_weather(
    month TEXT PRIMARY KEY,
    sunny FLOAT,
    part_cl FLOAT,
    full_cl FLOAT,
    precip FLOAT,
    city TEXT)
'''

cmd2 = '''CREATE TABLE IF NOT EXISTS
gilbert_weather(
    month TEXT PRIMARY KEY,
    sunny FLOAT,
    part_cl FLOAT,
    full_cl FLOAT,
    precip FLOAT,
    city TEXT)
'''

cmd3 = '''CREATE TABLE IF NOT EXISTS
tucson_weather(
    month TEXT PRIMARY KEY,
    sunny FLOAT,
    part_cl FLOAT,
    full_cl FLOAT,
    precip FLOAT,
    city TEXT)
'''

cur.execute(cmd)
cur.execute(cmd2)
cur.execute(cmd3)
```



```
for index, row in df_split[0].iterrows():
    cur.execute(f'''INSERT INTO flagstaff_weather VALUES (?,?,?,?,?,?)''',(
    row['month'],
    row['sunny'],
    row['part_cl'],
    row['full_cl'],
    row['precip'],
    row['city']
    ))
conn.commit()

for index, row in df_split[1].iterrows():
    cur.execute(f'''INSERT INTO gilbert_weather VALUES (?,?,?,?,?,?)''',(
    row['month'],
    row['sunny'],
    row['part_cl'],
    row['full_cl'],
    row['precip'],
    row['city']
    ))
conn.commit()

for index, row in df_split[2].iterrows():
    cur.execute(f'''INSERT INTO tucson_weather VALUES (?,?,?,?,?,?)''',(
    row['month'],
    row['sunny'],
    row['part_cl'],
    row['full_cl'],
    row['precip'],
    row['city']
    ))
conn.commit()
```



	month	sunny	part_cl	full_cl	precip	city
0	jan	0.45	0.23	0.32	0.13	Tucson
1	feb	0.45	0.24	0.31	0.14	Tucson
2	mar	0.48	0.23	0.29	0.12	Tucson
3	apr	0.55	0.23	0.21	0.07	Tucson
4	may	0.64	0.23	0.13	0.07	Tucson
5	jun	0.70	0.20	0.10	0.06	Tucson
6	jul	0.32	0.38	0.29	0.32	Tucson
7	aug	0.37	0.40	0.23	0.29	Tucson
8	sep	0.63	0.23	0.13	0.18	Tucson
9	oct	0.64	0.20	0.16	0.11	Tucson
10	nov	0.60	0.20	0.20	0.11	Tucson
11	dec	0.48	0.20	0.32	0.13	Tucson

Challenges Faced

- The programming needed throughout the project was tough and took some time
- Figuring out what type of plots would best represent the data

Limitations

- Time
- Time
- Time

Future Ideas

- Use more cities in AZ and around the US.
- Scrape more images and tables from the US City Data website.
- Find ways to make the code more efficient.