Analyzing COVID-19 Correlations

Case Question:

1. Based on age and gender, is there any correlation between the symptom status?

Research Method

1. **Plan Out the Goal**
   - Drew out imaginary graphs based on what we wanted to compare:
     - # of Asymptomatic/Symptomatic patients by age group and gender
     - % of Asymptomatic/Symptomatic patients by age group and gender

2. **Filter the Data / Data Wrangling**
   - Created a new dataframe based on the COVID-19 Cases Data which only included the columns:
     - Sex, age_group, and symptom_status
   - Removed unknown,none, and missing values
   - Counted and sorted the values

3. **Visualizations**
   - Transformed our data into 2 tables and 8 pie charts using Seaborn and Matplotlib

How Many People Were Tested?

Case Months : Jan 2020 - Dec 2021

- **<u>Symptom Status</u>**
  - Symptomatic: 227680
  - Asymptomatic: 7700
- **<u>Sex</u>**
  - Female :   123432
  - Male: 111948
- **<u>Age Groups (years)</u>**
  - 0 - 17:   40323
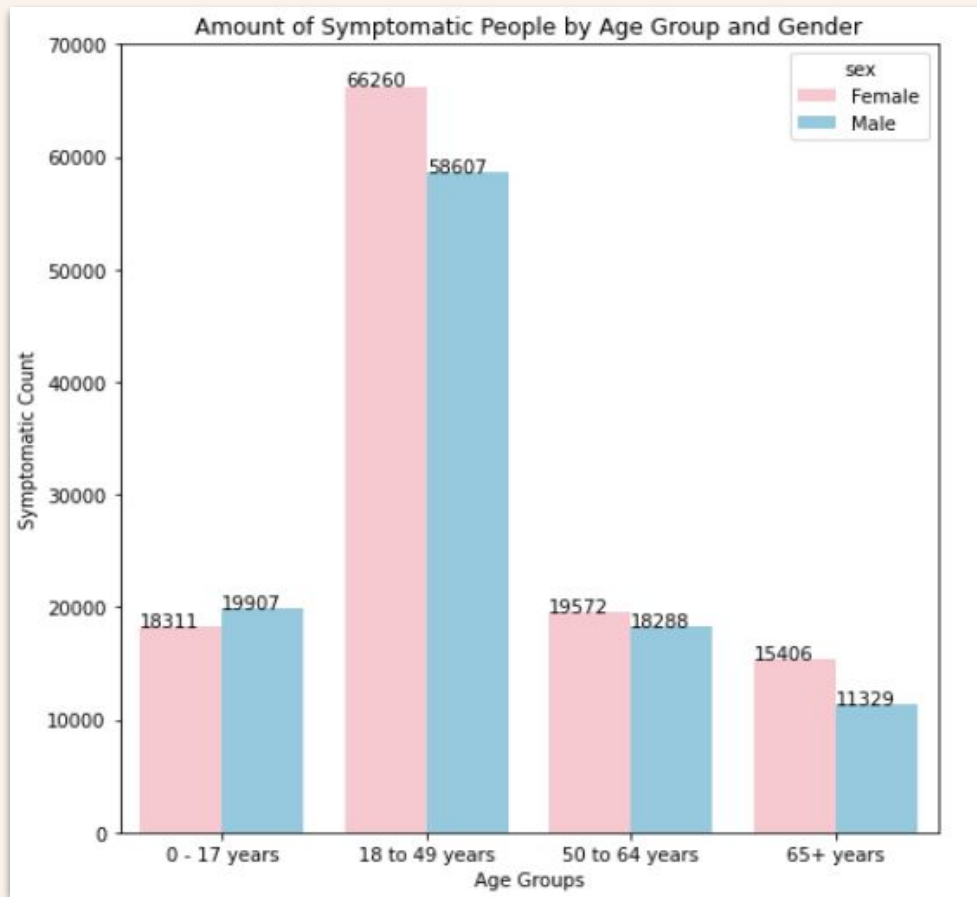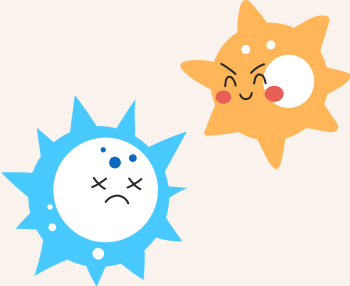 - 18 - 49 :   128297
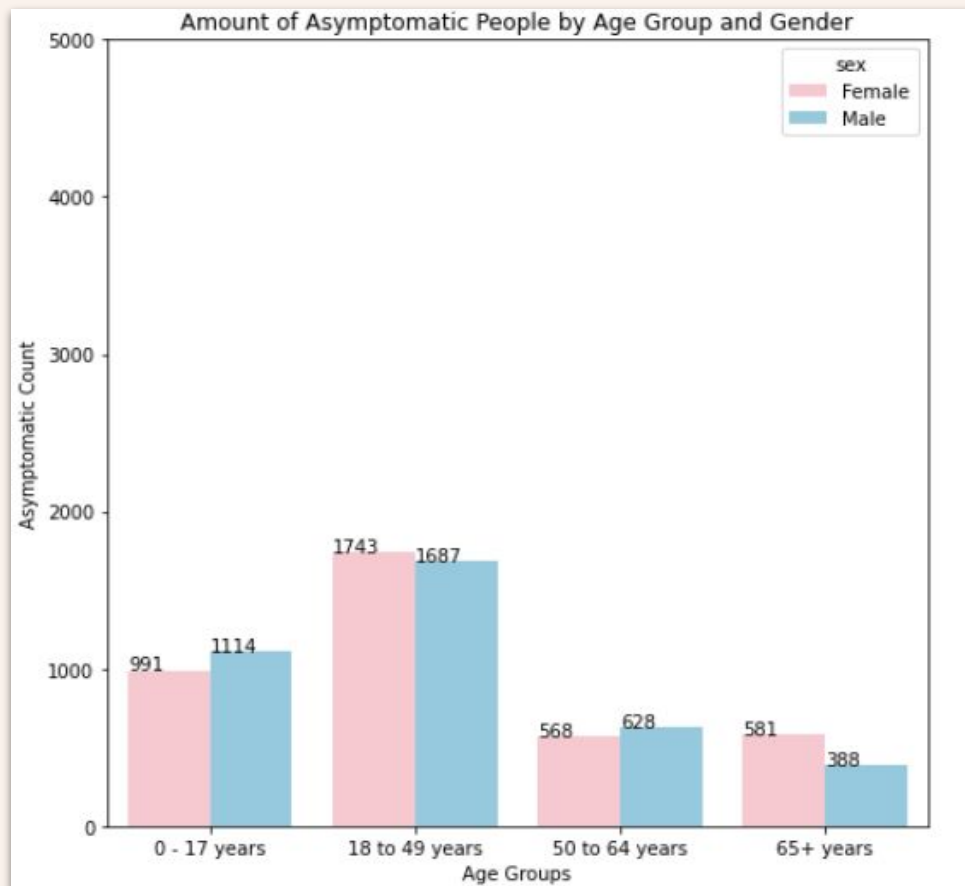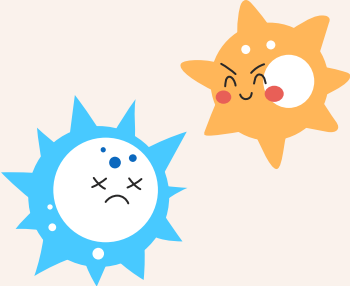 - 50 - 64 :   39056
  - 65+:   27704

Fig. 1.1

Fig. 1.2

Symptom Status by Age Group

Fig. 3.1

Fig. 3.2

% of Symptomatic Cases by Gender

Female 53%

Male 47%

% of Symptomatic Cases by Age Group

18 to 49 years 55%
0 - 17 years 17%
65+ years 12%
50 to 64 years 17%

Fig. 2.1

% of Asymptomatic Cases by Gender

Female 50%

Male 50%

% of Asymptomatic Cases by Age Group

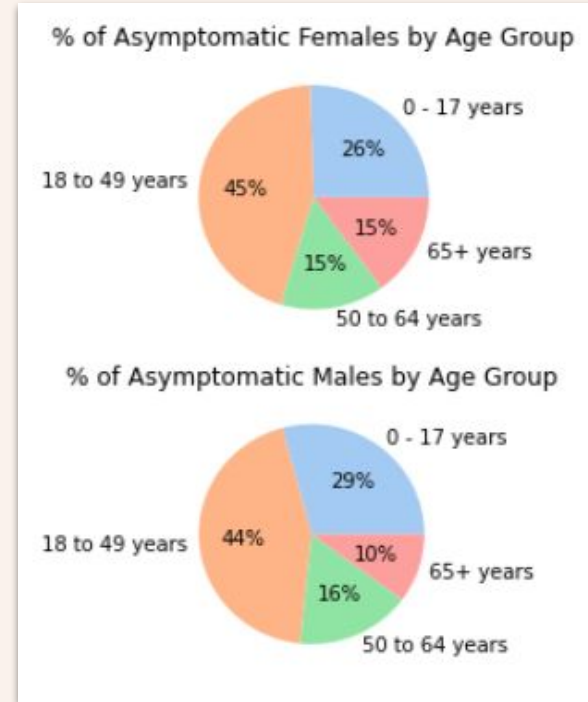18 to 49 years 45%
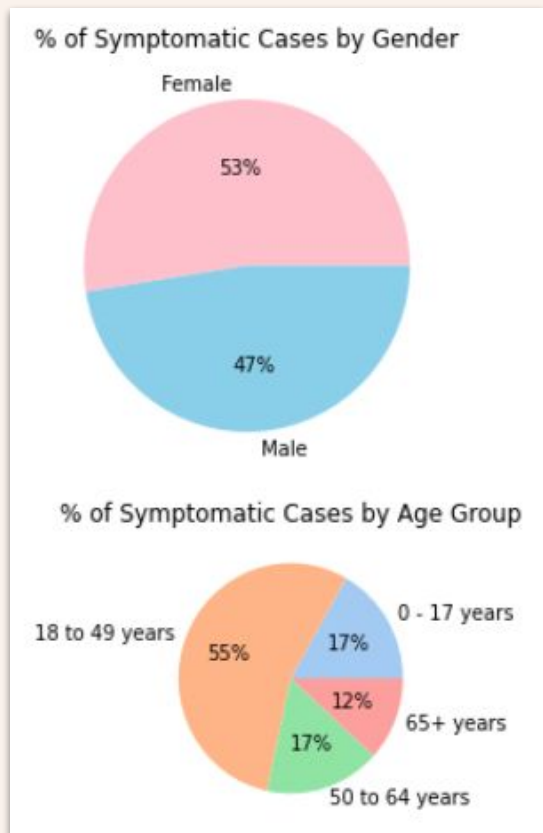0 - 17 years 27%
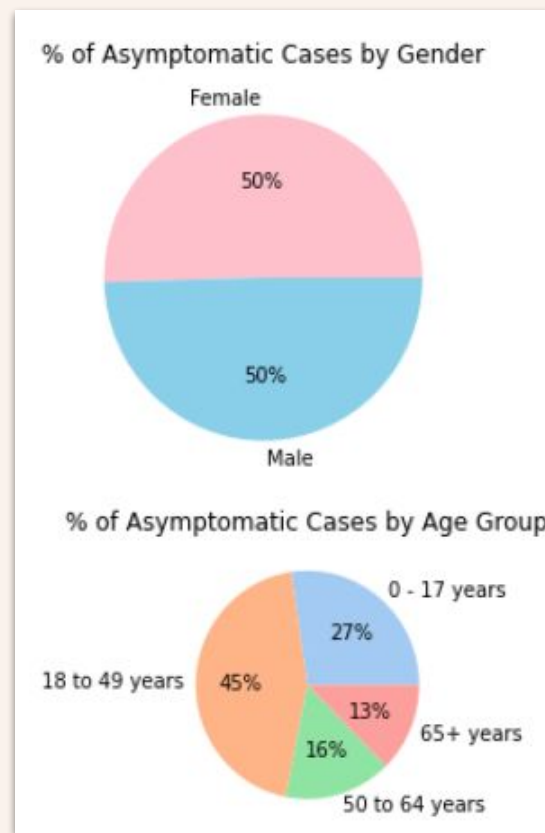65+ years 13%
50 to 64 years 16%
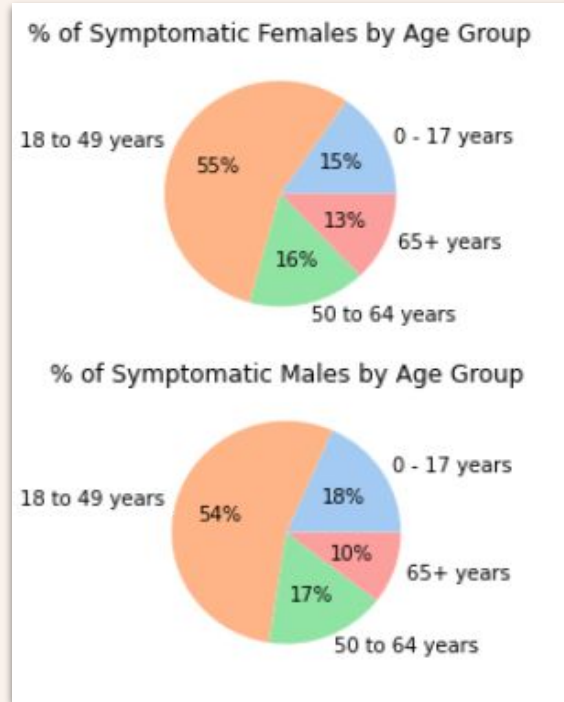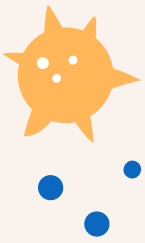
Fig. 2.2

Key Conclusions

1. **<u>Symptomatic cases are more common in females than males.</u>**
   - The difference in distribution is especially prominent in the 18-49 year old age group[1.1]

2. **<u>Asymptomatic cases are fairly distributed amongst female and male sex of all ages.</u>**[2.2]

3. **<u>18-49 year olds, regardless of gender, are more likely to catch COVID</u>**
   - This may be due to the uneven sampling distribution and large age range.
   - Largest generation in America (Millennials) is between 18-49 years old.