# IT350 : Data Analytics
# Lab Assignment 1

**Name:** Chinmayi C. Ramakrishna
**Roll No.:** 181IT113

Date of Submission : 24th Jan, 2021

---

**Dataset 1:**

| onth | day | cityid | gps_retail_and_recreation | gps_grocery_and_pharmacy | gps_parks | gps_transit_stations | gps_workplaces | gps_residential | gps_away_from_home |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 24 | 1 | 0.00571 | -0.00286 | .0714 | .00286 | 0.02140 | -0.00143 | 0.000625 |
| 2 | 24 | 2 | 0.02000 | -0.02410 | .139 | -.0206 | -0.03260 | 0.01160 | -0.009070 |
| 2 | 24 | 3 | 0.04000 | 0.02710 | .281 | .0329 | 0.02570 | -0.00857 | 0.014700 |
| 2 | 24 | 4 | 0.02140 | -0.00714 | .00286 | -.00286 | 0.02290 | 0.00286 | -0.002060 |
| 2 | 24 | 5 | 0.03290 | -0.00143 | .0386 | .0471 | 0.01000 | 0.00286 | -0.004610 |
| 2 | 24 | 6 | 0.00571 | -0.02000 | .0214 | .00143 | 0.01570 | 0.00143 | -0.003320 |
| 2 | 24 | 7 | 0.01140 | -0.01290 | .00857 | -.01 | 0.03140 | -0.00286 | 0.002660 |
| 2 | 24 | 8 | 0.02290 | -0.00571 | .0657 | .0271 | 0.00286 | 0.00286 | -0.005010 |
| 2 | 24 | 9 | 0.02430 | -0.01140 | .381 | .0386 | -0.03430 | 0.00000 | 0.001420 |
| 2 | 24 | 10 | 0.00857 | -0.00286 | -.0114 | .0371 | 0.02860 | -0.00286 | 0.003970 |

Fig 1. Google_Mobility_City_Daily table

**Details of the table content:**

| **Attribute** | Description |
|---|---|
| **gps_retail_and_recreation** | Time spent at retail and recreation locations. |
| **gps_grocery_and_pharmacy** | Time spent at grocery and pharmacy locations. |
| **gps_parks** | Time spent at parks. |
| **gps_transit_stations** | Time at inside transit stations. |
| **gps_workplaces** | Time spent at work places. |
| **gps_residential** | Time spent at residential locations. |
| **gps_away_from_home** | Time spent outside of residential locations. |

Table 1. Details of attributes of Google_Mobility_City_Daily

**Data details:**

The data in the table refers to the percentage ( in decimal) of time spent at various places in comparison to the baseline. Baseline refers to mobility before the pandemic outbreak.

**Frequency Distribution:**

```
gps_retail_and_recreation: 896
gps_grocery_and_pharmacy: 2323
gps_parks: 698
gps_transit_stations: 856
gps_workplaces: 320
gps_residential: 36
gps_away_from_home: 1040
```

Fig 2. Frequency distribution of Google_Mobility_City_Daily

According to the frequency distribution, the following can be analysed:

- 896 rows show time spent at retail and recreation places 20 % above the baseline.
- 2323 rows show time spent at grocery and pharmacy 20% above the baseline.
- 698 rows show time spent at parks 20% above the baseline.
- 856 rows show time spent at transit stations 20% below the baseline.
- 320 rows show time spent at workplaces 0.025 above the baseline.
- 36 rows show time spent at  residential places 0.3 above the baseline.
- 1040 rows show time spent away from home 0.02 below the baseline.

```
month
1      159
2      318
3     1643
4     1590
5     1643
6     1590
7     1643
8     1643
9     1590
10    1643
11    1590
12    1643
Name: month, dtype: int64
```

Fig 3. Data collected throughout the year

More data is collected towards the end of the year. It means there is greater mobility variations seen towards the end of the year.

**Graphical Representation:**
Gps_workplaces and gps_residential attributes have been used to plot in the histogram as they are important determinants of mobility during the pandemic.

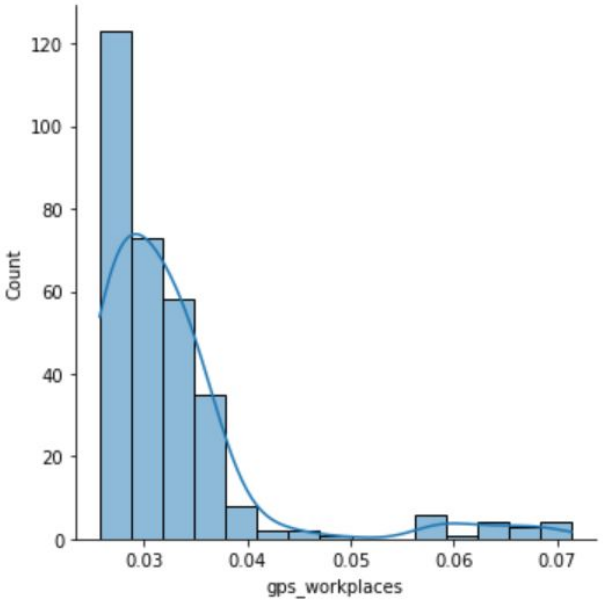


Fig 4. Count vs gps_workplaces

The above histogram shows the number of rows in the dataset with time spent at workplaces 0.025 above baseline.

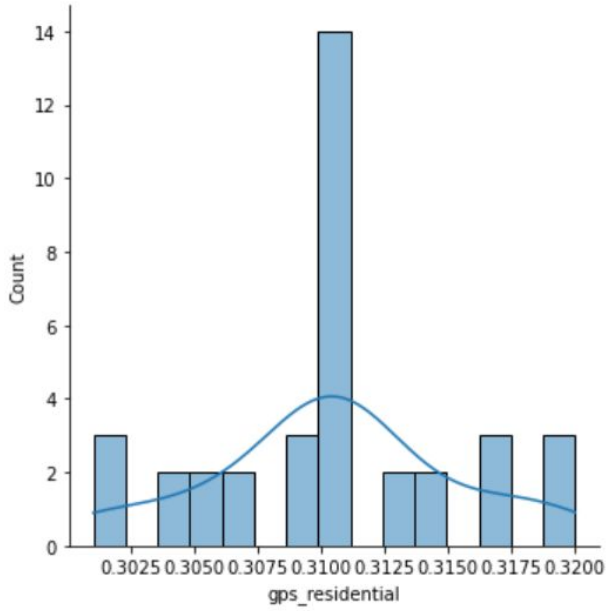


Fig 5. Count vs gps_residential

The above histogram shows the number of rows in the dataset with time spent at residential place 0.3 above the baseline.

**Summary Statistics:**

| | year | month | day | cityid | gps_retail_and_recreation | gps_grocery_and_pharmacy | gps_workplaces | gps_residential | gps_a |
|---|---|---|---|---|---|---|---|---|---|
| count | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | |
| mean | 2020.009524 | 7.333333 | 15.876190 | 27.000000 | -0.252719 | -0.080886 | -0.340571 | 0.114458 | |
| std | 0.097127 | 2.996384 | 8.938179 | 15.297517 | 0.157320 | 0.093898 | 0.129044 | 0.057170 | |
| min | 2020.000000 | 1.000000 | 1.000000 | 1.000000 | -0.764000 | -0.374000 | -0.711000 | -0.025700 | |
| 25% | 2020.000000 | 5.000000 | 8.000000 | 14.000000 | -0.353000 | -0.143000 | -0.413000 | 0.081400 | |
| 50% | 2020.000000 | 7.000000 | 16.000000 | 27.000000 | -0.233000 | -0.084300 | -0.344000 | 0.108000 | |
| 75% | 2020.000000 | 10.000000 | 24.000000 | 40.000000 | -0.160000 | -0.024300 | -0.291000 | 0.147000 | |
| max | 2021.000000 | 12.000000 | 31.000000 | 53.000000 | 0.481000 | 0.457000 | 0.071400 | 0.320000 | |

| year | month | day | cityid | gps_retail_and_recreation | gps_grocery_and_pharmacy | gps_workplaces | gps_residential | gps_away_from_home |
|---|---|---|---|---|---|---|---|---|
| 000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 | 16695.000000 |
| 009524 | 7.333333 | 15.876190 | 27.000000 | -0.252719 | -0.080886 | -0.340571 | 0.114458 | -0.134725 |
| 097127 | 2.996384 | 8.938179 | 15.297517 | 0.157320 | 0.093898 | 0.129044 | 0.057170 | 0.069940 |
| 000000 | 1.000000 | 1.000000 | 1.000000 | -0.764000 | -0.374000 | -0.711000 | -0.025700 | -0.391000 |
| 000000 | 5.000000 | 8.000000 | 14.000000 | -0.353000 | -0.143000 | -0.413000 | 0.081400 | -0.176000 |
| 000000 | 7.000000 | 16.000000 | 27.000000 | -0.233000 | -0.084300 | -0.344000 | 0.108000 | -0.126000 |
| 000000 | 10.000000 | 24.000000 | 40.000000 | -0.160000 | -0.024300 | -0.291000 | 0.147000 | -0.092500 |
| 000000 | 12.000000 | 31.000000 | 53.000000 | 0.481000 | 0.457000 | 0.071400 | 0.320000 | 0.035600 |

Fig 6. Summary statistics of Google_Mobility_City_Daily

From the summary statistics, we can conclude that there was an increase in the overall mobility to grocery and pharmaceutical stores. Recreational places and workplaces were visited less than the baseline. There was an increase in the time spent at home after the pandemic outbreak. Data from gps_workplaces deviates less from the mean of -0.34. A less deviation from mean is observed in the gps_residential attribute.

Graphical representation suggests that the dataset shows unimodal and bimodal shape for the two attributes chosen. The graph is positively skewed in Fig 4. and symmetrical in Fig 5.

**Dataset 2:**

| | country | region | region_code | start_date | end_date | year | week | population | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deat |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Spain | Andalusia | 1 | 2020-01-01 | 2020-01-07 | 2020 | 1 | 8405294 | 1542 | 0.0 | 1554.0 | -12.0 | 154 |
| 1 | Spain | Andalusia | 1 | 2020-01-08 | 2020-01-14 | 2020 | 2 | 8405294 | 1663 | 0.0 | 1646.0 | 17.0 | 166 |
| 2 | Spain | Andalusia | 1 | 2020-01-15 | 2020-01-21 | 2020 | 3 | 8405294 | 1812 | 0.0 | 1629.5 | 182.5 | 181 |
| 3 | Spain | Andalusia | 1 | 2020-01-22 | 2020-01-28 | 2020 | 4 | 8405294 | 1759 | 0.0 | 1656.0 | 103.0 | 175 |
| 4 | Spain | Andalusia | 1 | 2020-01-29 | 2020-02-04 | 2020 | 5 | 8405294 | 1796 | 0.0 | 1635.5 | 160.5 | 179 |
| 5 | Spain | Andalusia | 1 | 2020-02-05 | 2020-02-11 | 2020 | 6 | 8405294 | 1601 | 0.0 | 1646.5 | -45.5 | 160 |
| 6 | Spain | Andalusia | 1 | 2020-02-12 | 2020-02-18 | 2020 | 7 | 8405294 | 1512 | 0.0 | 1612.0 | -100.0 | 151 |
| 7 | Spain | Andalusia | 1 | 2020-02-19 | 2020-02-25 | 2020 | 8 | 8405294 | 1430 | 0.0 | 1572.5 | -142.5 | 143 |
| 8 | Spain | Andalusia | 1 | 2020-02-26 | 2020-03-03 | 2020 | 9 | 8405294 | 1393 | 0.0 | 1537.5 | -144.5 | 139 |
| 9 | Spain | Andalusia | 1 | 2020-03-04 | 2020-03-10 | 2020 | 10 | 8405294 | 1363 | 0.0 | 1477.0 | -114.0 | 136 |

| ion | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100k | excess_deaths_per_100k | excess_deaths_pct_change |
|---|---|---|---|---|---|---|---|---|
| 294 | 1542 | 0.0 | 1554.0 | -12.0 | 1542.0 | 0.0 | -0.142767 | -0.007722 |
| 294 | 1663 | 0.0 | 1646.0 | 17.0 | 1663.0 | 0.0 | 0.202253 | 0.010328 |
| 294 | 1812 | 0.0 | 1629.5 | 182.5 | 1812.0 | 0.0 | 2.171251 | 0.111998 |
| 294 | 1759 | 0.0 | 1656.0 | 103.0 | 1759.0 | 0.0 | 1.225418 | 0.062198 |
| 294 | 1796 | 0.0 | 1635.5 | 160.5 | 1796.0 | 0.0 | 1.909511 | 0.098135 |
| 294 | 1601 | 0.0 | 1646.5 | -45.5 | 1601.0 | 0.0 | -0.541326 | -0.027634 |
| 294 | 1512 | 0.0 | 1612.0 | -100.0 | 1512.0 | 0.0 | -1.189726 | -0.062035 |
| 294 | 1430 | 0.0 | 1572.5 | -142.5 | 1430.0 | 0.0 | -1.695360 | -0.090620 |
| 294 | 1393 | 0.0 | 1537.5 | -144.5 | 1393.0 | 0.0 | -1.719155 | -0.093984 |
| 294 | 1363 | 0.0 | 1477.0 | -114.0 | 1363.0 | 0.0 | -1.356288 | -0.077183 |

Fig 7. Spain_excess_deaths table

The table includes the attributes: country, region, region_code, start_date, end_date, year, week, population, total_deaths, covid_deaths, expected_deaths, excess_deaths, non_covid_deaths, covid_deaths_per_100k, excess_deaths_per_100k, excess_deaths_pct_change.

**Data details:**

Negative values in the table indicate its value below the baseline. Baseline is the number of deaths before the pandemic outbreak.

**Frequency Distribution:**

```
total_deaths: 98
covid_deaths: 11
expected_deaths: 63
excess_deaths: 193
non_covid_deaths: 934
covid_deaths_per_100k: 48
excess_deaths_per_100k: 726
excess_deaths_pct_change: 504
```

Fig 8. Frequency distribution for spain_excess_deaths

According to the frequency table, the following can be analysed:

- 98 rows in the dataset have total deaths greater than 1500.
- 11 rows in the dataset have covid deaths greater than 1500
- 63 rows in the dataset have expected deaths greater than 1500.
- 193 rows in the dataset have excess deaths greater than 100
- 934 rows in the dataset have non covid deaths greater than 20
- 48 rows in the dataset have a value of covid deaths per 100k, 10 greater than baseline.
- 726 rows in the dataset have excess deaths per 100k value greater than 0.1
- 504 rows in the dataset have excess deaths pct change greater than 0.1

**Graphical Representation:**
covid_deaths and expected_deaths have been taken as the attributes for histogram representation because it gives more insight on the pandemic situation. The two histograms can be compared to get some conclusion.

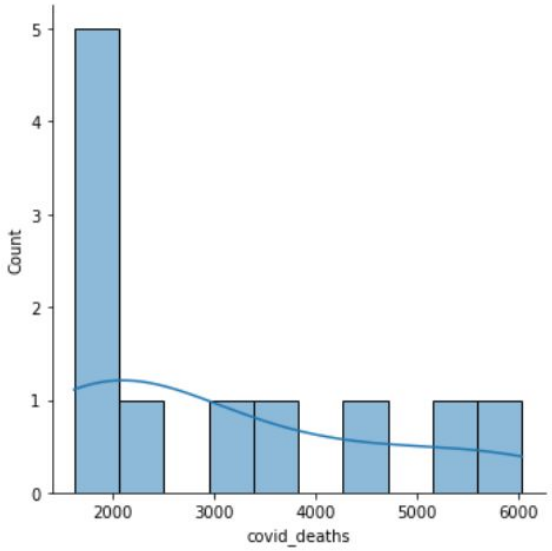


Fig 9. Count vs covid_deaths

The above histogram shows that most of the regions have deaths around 2000.

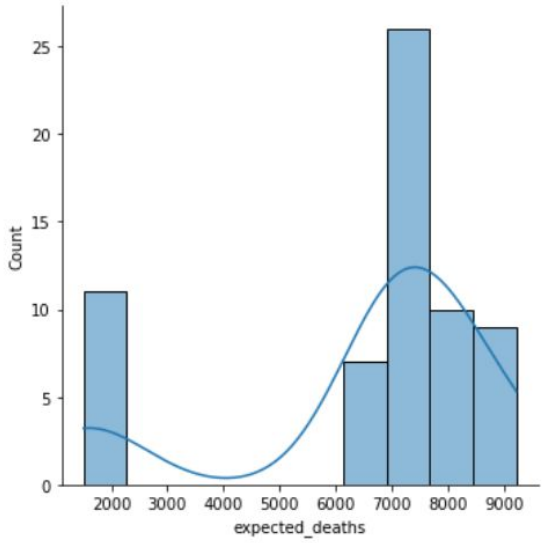


Fig 10. Count vs expected_deaths

The above histogram shows that deaths expected were higher than actual deaths.

**Summary Statistics:**

| | region_code | year | week | population | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100l |
|---|---|---|---|---|---|---|---|---|---|---|
| count | 1040.000000 | 1040.0 | 1040.000000 | 1.040000e+03 | 1040.000000 | 1023.000000 | 1040.000000 | 1040.000000 | 1023.000000 | 1023.00000( |
| mean | 9.500000 | 2020.0 | 26.500000 | 4.673304e+06 | 889.336538 | 97.704790 | 756.381010 | 132.955529 | 806.037146 | 2.11533( |
| std | 5.769056 | 0.0 | 15.015552 | 9.947887e+06 | 1987.983901 | 388.908205 | 1620.299624 | 663.504432 | 1732.153506 | 3.82860: |
| min | 0.000000 | 2020.0 | 1.000000 | 8.466700e+04 | 4.000000 | 0.000000 | 7.000000 | -619.000000 | 4.000000 | 0.00000( |
| 25% | 4.750000 | 2020.0 | 13.750000 | 9.298350e+05 | 140.750000 | 0.000000 | 136.500000 | -2.000000 | 142.000000 | 0.00000( |
| 50% | 9.500000 | 2020.0 | 26.500000 | 1.754952e+06 | 305.000000 | 8.000000 | 276.000000 | 15.500000 | 285.000000 | 0.58604: |
| 75% | 14.250000 | 2020.0 | 39.250000 | 3.261577e+06 | 759.250000 | 52.000000 | 693.875000 | 66.625000 | 700.500000 | 2.61520- |
| max | 19.000000 | 2020.0 | 52.000000 | 4.673304e+07 | 19358.000000 | 6032.000000 | 9224.500000 | 11499.000000 | 13326.000000 | 32.49939( |

| on | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100k | excess_deaths_per_100k | excess_deaths_pct_change |
|---|---|---|---|---|---|---|---|---|
| 03 | 1040.000000 | 1023.000000 | 1040.000000 | 1040.000000 | 1023.000000 | 1023.000000 | 1040.000000 | 1040.000000 |
| 06 | 889.336538 | 97.704790 | 756.381010 | 132.955529 | 806.037146 | 2.115330 | 2.909169 | 0.208928 |
| 06 | 1987.983901 | 388.908205 | 1620.299624 | 663.504432 | 1732.153506 | 3.828603 | 6.507072 | 0.459258 |
| 04 | 4.000000 | 0.000000 | 7.000000 | -619.000000 | 4.000000 | 0.000000 | -8.267684 | -0.500000 |
| 05 | 140.750000 | 0.000000 | 136.500000 | -2.000000 | 142.000000 | 0.000000 | -0.172216 | -0.010519 |
| 06 | 305.000000 | 8.000000 | 276.000000 | 15.500000 | 285.000000 | 0.586047 | 1.423728 | 0.094987 |
| 06 | 759.250000 | 52.000000 | 693.875000 | 66.625000 | 700.500000 | 2.615204 | 3.675652 | 0.240518 |
| 07 | 19358.000000 | 6032.000000 | 9224.500000 | 11499.000000 | 13326.000000 | 32.499396 | 74.293126 | 4.763158 |

Fig 11. Summary statistics for spain_excess_deaths

From the above summary, we can conclude that mean total mortality was near to the mean expected deaths. Non covid deaths mainly contributed to the total deaths. Covid deaths per 100k remains low. Expected deaths attribute has a large deviation from its mean of 756. The expected deaths data is observed to be either much lower than mean or much higher. Excess deaths has a minimum of 619 deaths less than the baseline. Covid deaths per 100k remains pretty low in Spain. Covid doesn't seem to have affected much in the percentage death change.

The graph in Fig 10. is  unimodal and negatively skewed.

**Dataset 3:**

| | country | region | region_code | start_date | end_date | year | week | population | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_death |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | United States | Alabama | AL | 2019-12-29 | 2020-01-04 | 2020 | 1 | 4903185 | 1081 | 0 | 1167.309220 | -86.309220 | 108 |
| 1 | United States | Alabama | AL | 2020-01-05 | 2020-01-11 | 2020 | 2 | 4903185 | 1127 | 0 | 1195.142553 | -68.142553 | 112 |
| 2 | United States | Alabama | AL | 2020-01-12 | 2020-01-18 | 2020 | 3 | 4903185 | 1039 | 0 | 1153.142553 | -114.142553 | 103 |
| 3 | United States | Alabama | AL | 2020-01-19 | 2020-01-25 | 2020 | 4 | 4903185 | 1054 | 0 | 1144.975886 | -90.975886 | 105 |
| 4 | United States | Alabama | AL | 2020-01-26 | 2020-02-01 | 2020 | 5 | 4903185 | 1025 | 0 | 1140.142553 | -115.142553 | 102 |
| 5 | United States | Alabama | AL | 2020-02-02 | 2020-02-08 | 2020 | 6 | 4903185 | 1118 | 0 | 1155.475887 | -37.475887 | 111 |
| 6 | United States | Alabama | AL | 2020-02-09 | 2020-02-15 | 2020 | 7 | 4903185 | 1094 | 0 | 1142.309220 | -48.309220 | 109 |
| 7 | United States | Alabama | AL | 2020-02-16 | 2020-02-22 | 2020 | 8 | 4903185 | 1097 | 0 | 1128.309220 | -31.309220 | 109 |
| 8 | United States | Alabama | AL | 2020-02-23 | 2020-02-29 | 2020 | 9 | 4903185 | 1161 | 0 | 1132.642553 | 28.357447 | 116 |
| 9 | United States | Alabama | AL | 2020-03-01 | 2020-03-07 | 2020 | 10 | 4903185 | 1053 | 0 | 1125.371064 | -72.371064 | 105 |

| ion | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100k | excess_deaths_per_100k | excess_deaths_pct_change |
|---|---|---|---|---|---|---|---|---|
| 185 | 1081 | 0 | 1167.309220 | -86.309220 | 1081 | 0.0 | -1.760268 | -0.073939 |
| 185 | 1127 | 0 | 1195.142553 | -68.142553 | 1127 | 0.0 | -1.389761 | -0.057016 |
| 185 | 1039 | 0 | 1153.142553 | -114.142553 | 1039 | 0.0 | -2.327927 | -0.098984 |
| 185 | 1054 | 0 | 1144.975886 | -90.975886 | 1054 | 0.0 | -1.855445 | -0.079457 |
| 185 | 1025 | 0 | 1140.142553 | -115.142553 | 1025 | 0.0 | -2.348322 | -0.100990 |
| 185 | 1118 | 0 | 1155.475887 | -37.475887 | 1118 | 0.0 | -0.764317 | -0.032433 |
| 185 | 1094 | 0 | 1142.309220 | -48.309220 | 1094 | 0.0 | -0.985262 | -0.042291 |
| 185 | 1097 | 0 | 1128.309220 | -31.309220 | 1097 | 0.0 | -0.638549 | -0.027749 |
| 185 | 1161 | 0 | 1132.642553 | 28.357447 | 1161 | 0.0 | 0.578347 | 0.025037 |
| 185 | 1053 | 0 | 1125.371064 | -72.371064 | 1053 | 0.0 | -1.476001 | -0.064309 |

Fig 12. United_states_excess_deaths table

The table includes the attributes: country, region, region_code, start_date, end_date, year, week, population, total_deaths, covid_deaths, expected_deaths, excess_deaths, non_covid_deaths, covid_deaths_per_100k, excess_deaths_per_100k, excess_deaths_pct_change.

**Data details:**

Negative values in the table indicate its value below the baseline. Baseline is the number of deaths before the pandemic outbreak.

**Frequency Distribution:**

```
total_deaths: 683
covid_deaths: 55
expected_deaths: 566
excess_deaths: 992
non_covid_deaths: 2636
covid_deaths_per_100k: 52
excess_deaths_per_100k: 1981
excess_deaths_pct_change: 1268
```

Fig 13. Frequency distribution for united_states_excess_deaths

According to the frequency table, the following can be analysed:

- 683 rows in the dataset have total deaths greater than 1500.
- 55 rows in the dataset have covid deaths greater than 1500
- 566 rows in the dataset have expected deaths greater than 1500.
- 992 rows in the dataset have excess deaths greater than 100
- 2636 rows in the dataset have non covid deaths greater than 20
- 52 rows in the dataset have a value of covid deaths per 100k, 10 greater than baseline.
- 1981 rows in the dataset have excess deaths per 100k value greater than 0.1
- 1268 rows in the dataset have excess deaths pct change greater than 0.1

The frequency table comparisons of spain_excess_deaths and united_states_excess_deaths, can give the following conclusions:

- The United states has higher covid deaths per 100k than Spain
- Number of Excess deaths percentage change also is higher for the United States.

**Graphical Representations:**
covid_deaths and expected_deaths have been taken as the attributes for histogram representation because it gives more insight on the pandemic situation.

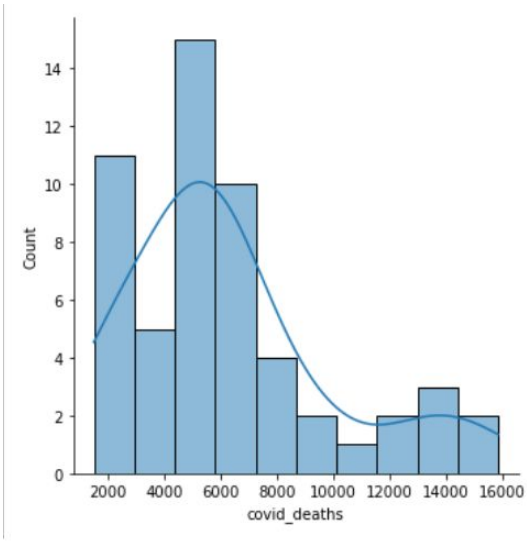


Fig 14. Count vs covid_deaths

The above histogram shows that there are more regions with death recorded around 6000.

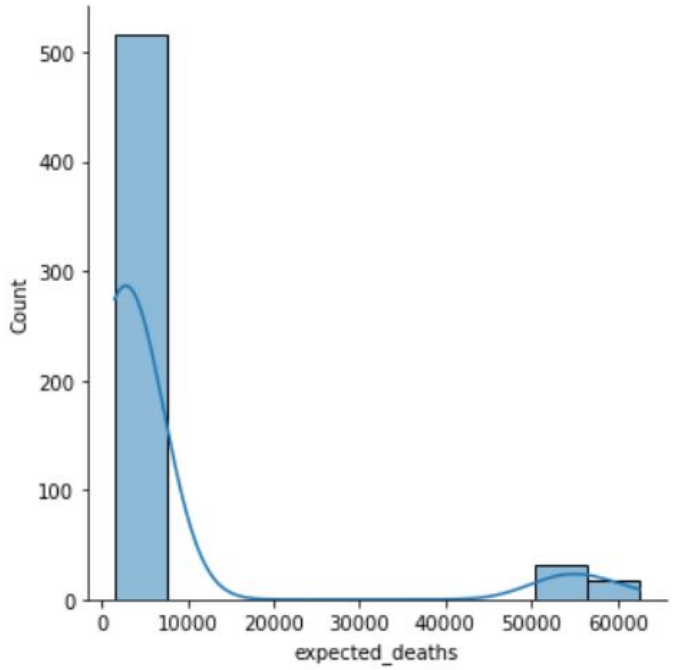


Fig 15. Count vs expected_deaths

The above histogram shows a higher expectation of deaths around 9000.

**Summary Statistics:**

| | year | week | population | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100k | excess_dea |
|---|---|---|---|---|---|---|---|---|---|---|
| count | 2640.0 | 2640.000000 | 2.640000e+03 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | |
| mean | 2020.0 | 25.420833 | 1.252048e+07 | 2409.587500 | 229.551515 | 2115.095587 | 294.491913 | 2180.035985 | 1.759593 | |
| std | 0.0 | 14.402374 | 4.419981e+07 | 8522.939606 | 1044.282976 | 7505.935589 | 1349.503969 | 7714.008647 | 3.272766 | |
| min | 2020.0 | 1.000000 | 5.787590e+05 | 61.000000 | -5.000000 | 77.022979 | -1987.689362 | -646.000000 | -0.471983 | |
| 25% | 2020.0 | 13.000000 | 1.792147e+06 | 344.000000 | 1.000000 | 311.892766 | 3.791702 | 318.000000 | 0.011716 | |
| 50% | 2020.0 | 25.000000 | 4.648794e+06 | 954.000000 | 33.000000 | 864.375745 | 50.994894 | 871.500000 | 0.878822 | |
| 75% | 2020.0 | 38.000000 | 8.398748e+06 | 1532.500000 | 121.000000 | 1366.163191 | 172.051489 | 1404.250000 | 2.059243 | |
| max | 2020.0 | 50.000000 | 3.283005e+08 | 79053.000000 | 15851.000000 | 62533.689362 | 22464.172766 | 65535.000000 | 63.283242 | |

| ɔn | total_deaths | covid_deaths | expected_deaths | excess_deaths | non_covid_deaths | covid_deaths_per_100k | excess_deaths_per_100k | excess_deaths_pct_change |
|---|---|---|---|---|---|---|---|---|
| 03 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 | 2640.000000 |
| 07 | 2409.587500 | 229.551515 | 2115.095587 | 294.491913 | 2180.035985 | 1.759593 | 2.314486 | 0.140070 |
| 07 | 8522.939606 | 1044.282976 | 7505.935589 | 1349.503969 | 7714.008647 | 3.272766 | 4.385713 | 0.291478 |
| 05 | 61.000000 | -5.000000 | 77.022979 | -1987.689362 | -646.000000 | -0.471983 | -21.039070 | -0.796012 |
| 06 | 344.000000 | 1.000000 | 311.892766 | 3.791702 | 318.000000 | 0.011716 | 0.110806 | 0.006861 |
| 06 | 954.000000 | 33.000000 | 864.375745 | 50.994894 | 871.500000 | 0.878822 | 1.580830 | 0.094208 |
| 06 | 1532.500000 | 121.000000 | 1366.163191 | 172.051489 | 1404.250000 | 2.059243 | 3.278029 | 0.193606 |
| 08 | 79053.000000 | 15851.000000 | 62533.689362 | 22464.172766 | 65535.000000 | 63.283242 | 80.863025 | 6.355981 |

Fig 16. Summary statistics of us_excess_deaths

The above summary can help us draw some conclusions. Total number of deaths has a large deviation from its mean. Expected deaths attribute has the greatest deviation from its mean. The prediction of expected deaths is quite distorted. Number of covid deaths has a low mean but the rest of the data is quite deviated from the mean. Deviation from the mean is large for most data fields. There is a subsequent amount of covid deaths. There is a little percentage change to the deaths due to covid.

Both the graphs are positively skewed and unimodal.