Stage2_Task3_Deaths

March 14, 2023

0.0.1 Identify 3 counties within a state of your choice with high death rates.

[1]:
```
import pandas as pd
import numpy as np
import statistics
```

[2]:
```
#I have selected the Albama State
selected_state = "AL"
#Readind the deaths data
deaths = pd.read_csv("../data/covid_deaths_usafacts.csv")
deaths.head()
```

[2]:

	countyFIPS	County Name	State	StateFIPS	2020-01-22	2020-01-23	\
0	0	Statewide Unallocated	AL	1	0	0	
1	1001	Autauga County	AL	1	0	0	
2	1003	Baldwin County	AL	1	0	0	
3	1005	Barbour County	AL	1	0	0	
4	1007	Bibb County	AL	1	0	0	

	2020-01-24	2020-01-25	2020-01-26	2020-01-27	…	2023-01-07	\
0	0	0	0	0	…	0	
1	0	0	0	0	…	230	
2	0	0	0	0	…	719	
3	0	0	0	0	…	103	
4	0	0	0	0	…	108	

	2023-01-08	2023-01-09	2023-01-10	2023-01-11	2023-01-12	2023-01-13	\
0	0	0	0	0	0	0	
1	230	230	230	230	230	230	
2	719	719	719	719	721	721	
3	103	103	103	103	103	103	
4	108	108	108	108	108	108	

	2023-01-14	2023-01-15	2023-01-16
0	0	0	0
1	230	230	230
2	721	721	721
3	103	103	103

```
4         108         108         108

[5 rows x 1095 columns]
```

[3]:
```
# using the melt function so that we get the all the dates in one column and␣
 ↪merging will be easy with enrichment data.
deaths_transpose = pd.melt(frame= deaths, id_vars=('countyFIPS','County␣
 ↪Name','State','StateFIPS'), var_name=["Date"], value_name='Number of Deaths')
deaths_transpose = deaths_transpose[deaths_transpose['countyFIPS'] != 0]
deaths_transpose.head()
```

[3]:

	countyFIPS	County Name	State	StateFIPS	Date	Number of Deaths
1	1001	Autauga County	AL	1	2020-01-22	0
2	1003	Baldwin County	AL	1	2020-01-22	0
3	1005	Barbour County	AL	1	2020-01-22	0
4	1007	Bibb County	AL	1	2020-01-22	0
5	1009	Blount County	AL	1	2020-01-22	0

[4]:
```
#Dropping the unwanted columns.
deaths_selected_state = deaths_transpose[deaths_transpose["State"] ==␣
 ↪selected_state]
deaths_selected_state = deaths_selected_state.drop(['countyFIPS', 'State',␣
 ↪'StateFIPS'], axis=1)
deaths_selected_state
```

[4]:

	County Name	Date	Number of Deaths
1	Autauga County	2020-01-22	0
2	Baldwin County	2020-01-22	0
3	Barbour County	2020-01-22	0
4	Bibb County	2020-01-22	0
5	Blount County	2020-01-22	0
…	…	…	…
3480433	Tuscaloosa County	2023-01-16	847
3480434	Walker County	2023-01-16	476
3480435	Washington County	2023-01-16	61
3480436	Wilcox County	2023-01-16	48
3480437	Winston County	2023-01-16	140

[73097 rows x 3 columns]

[5]:
```
#Number of new deaths daily in every county of selected state.
new_deaths_selected_state = deaths_selected_state
new_deaths_selected_state['Number of Deaths'] = new_deaths_selected_state.
 ↪groupby('County Name')['Number of Deaths'].diff()
new_deaths_selected_state
```

[5]:

	County Name	Date	Number of Deaths
1	Autauga County	2020-01-22	NaN
2	Baldwin County	2020-01-22	NaN
3	Barbour County	2020-01-22	NaN
4	Bibb County	2020-01-22	NaN
5	Blount County	2020-01-22	NaN
…	…	…	…
3480433	Tuscaloosa County	2023-01-16	0.0
3480434	Walker County	2023-01-16	0.0
3480435	Washington County	2023-01-16	0.0
3480436	Wilcox County	2023-01-16	0.0
3480437	Winston County	2023-01-16	0.0

[73097 rows x 3 columns]

[6]:
```
#Summing up the cases in every county and finding the three counties which have␣
 ↪higher number of deaths.
total_deaths_selected_state = new_deaths_selected_state
total_deaths_selected_state = total_deaths_selected_state.groupby('County␣
 ↪Name')['Number of Deaths'].sum().reset_index()
total_deaths_selected_state = total_deaths_selected_state.sort_values(by =␣
 ↪['Number of Deaths'], ascending=False).reset_index(drop=True)
total_deaths_selected_state.head(3)
```

[6]:

	County Name	Number of Deaths
0	Jefferson County	2495.0
1	Mobile County	1765.0
2	Madison County	1052.0