DATA WRANGLING PROJECT

How health problems in the United States are becoming more prevalent as a result of the rise in fast-food restaurants?

Group-2:

Sanjana Bhupathiraju
Priyanka Bathula
Bharath Simha Muthyala
Anandita Maurya

CONTENTS

Introduction…...……...…….2

Data Sources 3

Information Quality 4

Data Wrangling Process 4

Analysis & Results……..….10

External Materials…….…..11

Future Potential Data & Analysis.....……..…13

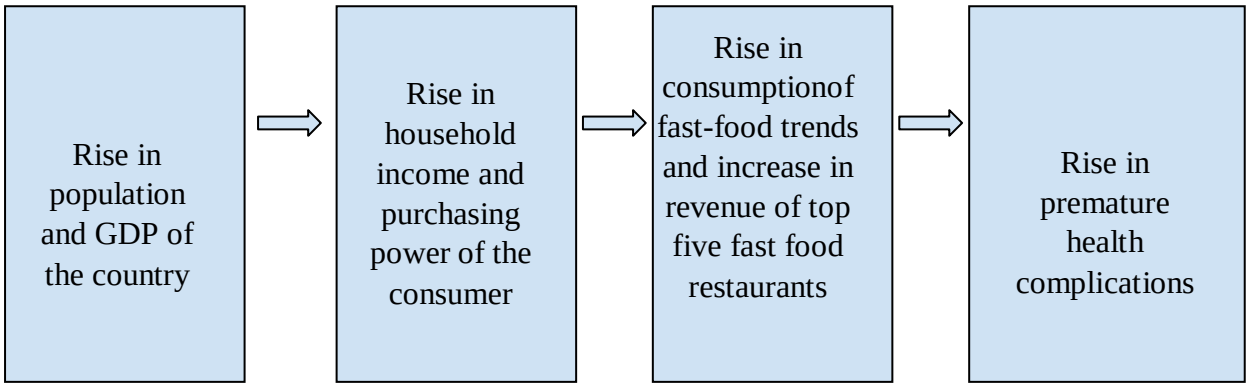
"People are fed by the food industry, which pays no attention to health and are treated by the health industry, which pays no attention to food."

In this busy world, it is very easy to grab food outside and finish a meal rather than making a meal at home. Over the years, the quick serving restaurants have become easily available at consumer's demand in every corner of the world. Eventually, the American Food and Beverage industry served as an inspiration for this project.

Given the convenience, it is easier to surrender control over the nutritional value of food. It is not the end of the world to eat fast food occasionally but consuming excessive amounts of fast food increases the likelihood of health issues including cardiovascular diseases and obesity. The excessive levels of fat, salt, and calories in fast food has earned a bad reputation among health-conscious diners. These foods lead to obesity, hypertension, and cholesterol elevation. These elements are well-known for the elevated danger of deteriorating the health. Therefore, this project aims to analyze - "***How health problems in the United States are becoming more prevalent as a result of the rise in fast food restaurants?"***

In order to conduct this analysis, the quarterly revenue data from 2015 through 2022 forthe following five publicly traded US companies: Starbucks, Domino's, Chipotle Mexican Grill, McDonald Corp, and Wendy's have been compiled. By comparing the five companies' financial results to four independent variables—GDP, total population estimates in the US, per capita income andpersonal consumption index—this project aims to highlight the –

- Increase in fast food restaurant's revenue as a result of rise in country's GDP
- Rise in health complications correlating to rise in standard of living of the population



The foremost and primary subject addressed in the project is the excessive growth in the consumption of fast food and processed food leading to deterioration of health and various

health complications. Customers have a greater tendency to shift towards a healthier lifestyle by eating a healthier diet. Thus, the food and beverage manufacturers must adjust their focus from categories and products to customers if they want to keep up with the expansion of this sector. This study also drives the government to take precautions and pass regulations to migrate industries and public towards a healthy eating environment.

Data sources:

The data sets are sourced from various government and private websites.

- The datasets about heart diseases and diabetes in United States has been derived from a government website, Centre for Disease Control and prevention where the number of reported cardiovascular disease cases are found from 2015-2021.

Website URL: https://www.cdc.gov/nchs/fastats/heart-disease.htm

- The data related to each fast-food restaurant (Wendy's, Starbucks, Chipotle, McDonalds and Domino's Pizza) is derived from a website named investing pro where the company's individual sales and revenue data was collected.

Website URL: https://www.investing.com/pro/watchlist/w-1606272.iwl/v-fecbbf0d/

- The data set for quarterly GDP of each state in the United States is found from the year 2012 to 2022. However, we couldn't derive GDP of fast-food industry alone. The total industry GDP is then taken into consideration.

Website URL: https://www.statista.com/statistics/188185/percent-change-from-preceding-period-in-real-gdp- in-the-us/

- The data set of the population census of each state and personal consumption of food and beverages from 2015 to 2021is collected from a government website, *census.gov*.

Website URL: https://data.census.gov/cedsci/

- The data of income per capita is obtained from a government website, Bureau of Economic Analysis.

Website URL: https://www.bea.gov/

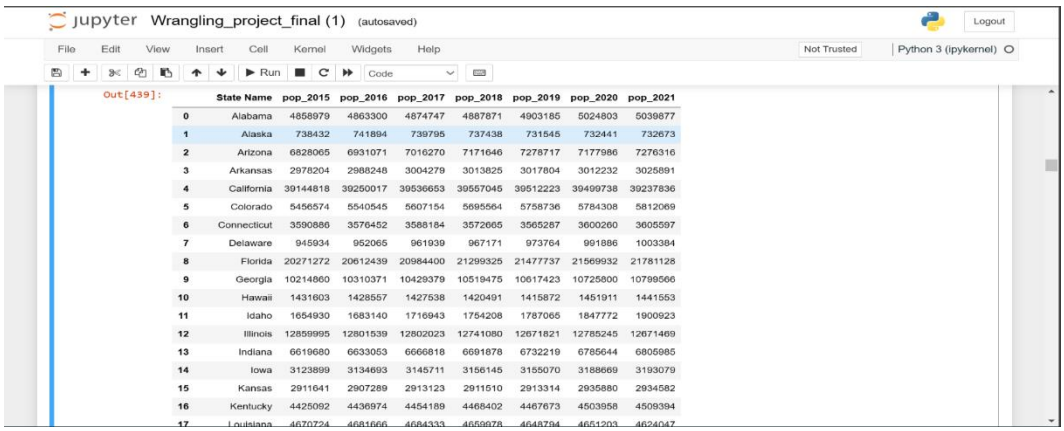
Information quality:

We are very certain that this data was sufficient to conduct our analysis through various wrangling techniques such as sorting, extracting, and evaluating the right information. However, the data collected required some formatting, profiling and preprocessing to obtain the right quality information. To maintain consistency in the data, the records of other states' data is excluded. As the data of cardiovascular diseases is not found for the years 2020 and 2021, the timeliness of the data is maintained by predicting the values for these years. The required amount of data is only extracted to avoid barriers to data accessibility. A few unnecessary columns with missing data were removed while concatenating a few elements of the data using data wrangling techniques. The data sets collected helped us in analyzing the data using data-wrangling tools and techniques to understand the relationship between the independent variable (F & B industry) and dependent variables (health complications).

Data Wrangling Process:

1. All the population files of each year from 2015 to 2021 were needed for our analysis but all the datasets had different format, attributes and datatypes. To format the data, all the csv files have been read using pd.read function. A subset is created for all the years with only required columns and concatenated based on the state names. It is observed that the data type of population was in float with decimals but as population cannot be decimals, used astype(int) function to convert the data type into integer. To improve the consistency of the data, used .isna().sum() function to look for null values. Later using the drop function, removed the null and inconsistent data that are not required for our analysis.

Image 1



	State Name	pop_2015	pop_2016	pop_2017	pop_2018	pop_2019	pop_2020	pop_2021
0	Alabama	4858979	4863300	4874747	4887871	4903185	5024803	5039877
1	Alaska	738432	741894	739795	737438	731545	732441	732673
2	Arizona	6828065	6931071	7016270	7171646	7278717	7177986	7276316
3	Arkansas	2978204	2988248	3004279	3013825	3017804	3012232	3025891
4	California	39144818	39250017	39536653	39557045	39512223	39499738	39237836
5	Colorado	5456574	5540545	5607154	5695564	5758736	5784308	5812069
6	Connecticut	3590886	3576452	3588184	3572665	3565287	3600260	3605597
7	Delaware	945934	952065	961939	967171	973764	991886	1003384
8	Florida	20271272	20612439	20984400	21299325	21477737	21569932	21781128
9	Georgia	10214860	10310371	10429379	10519475	10617423	10725800	10799566
10	Hawaii	1431603	1428557	1427538	1420491	1415872	1451911	1441553
11	Idaho	1654930	1683140	1716943	1754208	1787065	1847772	1900923
12	Illinois	12859995	12801539	12802023	12741080	12671821	12785245	12671469
13	Indiana	6619680	6633053	6666818	6691878	6732219	6785644	6805985
14	Iowa	3123899	3134693	3145711	3156145	3155070	3188669	3193079
15	Kansas	2911641	2907289	2913123	2911510	2913314	2935880	2934582
16	Kentucky	4425092	4436974	4454189	4468402	4467673	4503958	4509394
17	Louisiana	4670724	4681666	4684333	4659978	4648794	4651203	4624047

Image 1 is the final data set obtained after applying all the data wrangling steps on population datasets.

2. A dataset with all the GDP details from the year 2005 to 2022 has been found. To clean and wrangle the obtained data based on our requirements we used python. The total industry GDP of each state has been read where the data was given on a quarterly basis. The Annual GDP is extracted from the data as the analysis is to be done annually. As the analysis is to be done only for the 50 US states, all the other inconsistent data is dropped. To improve the consistency of the data, we checked for any null values and made sure that all the columns are of the same datatypes.

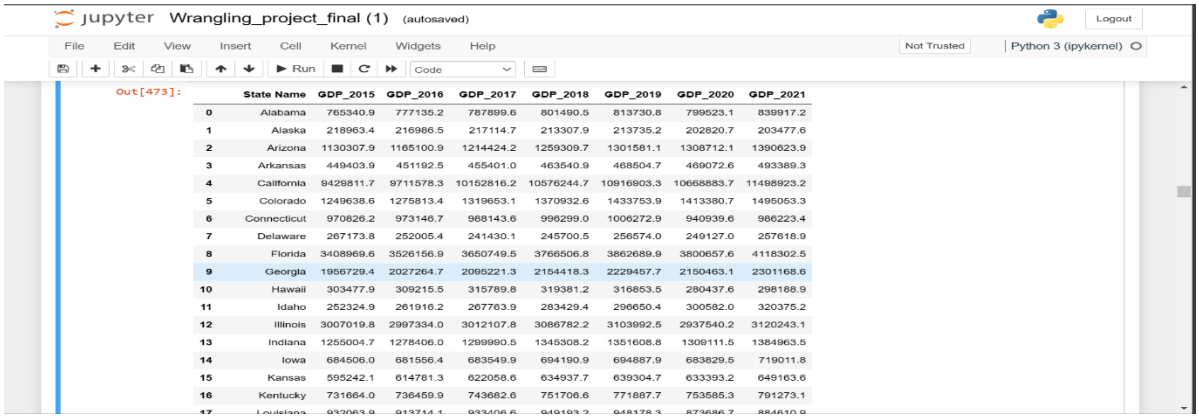
Image 2

```
In [123]: #Check for null values
          GDP_final.isna().sum()

Out[123]: GeoName     0
          2015        0
          2016        0
          2017        0
          2018        0
          2019        0
          2020        0
          2021        0
          dtype: int64
```

Image 2 is the code for finding null values in our GDP dataset.

Image 3



	State Name	GDP_2015	GDP_2016	GDP_2017	GDP_2018	GDP_2019	GDP_2020	GDP_2021
0	Alabama	765340.9	777135.2	787899.6	801490.5	813730.8	799523.1	839917.2
1	Alaska	218963.4	216986.5	217114.7	213307.9	213735.2	202820.7	203477.6
2	Arizona	1130307.9	1165100.9	1214424.2	1259309.7	1301581.1	1308712.1	1390623.9
3	Arkansas	449403.9	451192.5	455401.0	463540.9	468504.7	469072.6	493389.3
4	California	9429811.7	9711578.3	10152816.2	10576244.7	10916903.3	10668883.7	11498923.2
5	Colorado	1249638.6	1275813.4	1319653.1	1370932.6	1433753.9	1413380.7	1495053.3
6	Connecticut	970826.2	973146.7	988143.6	996299.0	1006272.9	940939.6	986223.4
7	Delaware	267173.8	252005.4	241430.1	245700.5	256574.0	249127.0	257618.9
8	Florida	3408969.6	3526156.9	3650749.5	3766506.8	3862689.9	3800657.6	4118302.5
9	Georgia	1956729.4	2027264.7	2095221.3	2154418.3	2229457.7	2150463.1	2301168.6
10	Hawaii	303477.9	309215.5	315789.8	319381.2	316853.5	280437.6	298188.9
11	Idaho	252324.9	261916.2	267763.9	283429.4	296650.4	300582.0	320375.2
12	Illinois	3007019.8	2997334.0	3012107.8	3086782.2	3103992.5	2937540.2	3120243.1
13	Indiana	1255004.7	1278406.0	1299990.5	1345308.2	1351608.8	1309111.5	1384963.5
14	Iowa	684506.0	681556.4	683549.9	694190.9	694887.9	683829.5	719011.8
15	Kansas	595242.1	614781.3	622058.6	634937.7	639304.7	633393.2	649163.6
16	Kentucky	731664.0	736459.9	743682.6	751706.6	771887.7	753585.3	791273.1

Image 3 is the final data set obtained after applying all the data wrangling steps on population datasets.

3. The Income statements of Wendy's, Starbucks, Chipotle, McDonalds, and Domino's have been found for the years 2015 to 2021. From which we have extracted the revenue of all the fast-food restaurants and converted into data frame. For better analysis of the data, we have converted the datatype of the columns from float to integer as different columns had different datatypes. The same data wrangling techniques are followed for all the five restaurants i.e., Wendy's, Starbucks, Chipotle, McDonalds, and Domino's Pizza. The final data of these restaurants is concatenated into one data frame based on the year.

Image 4

```
In [6]: #Considering only the Revenue of the fast food chain for analysis
        wendys = wendys[wendys['Unnamed: 0'] == 'Revenue (Reported)']
        wendys
```

Out[6]:

	Unnamed: 0	LTM.FQ-39	LTM.FQ-38	LTM.FQ-37	LTM.FQ-36	LTM.FQ-35	LTM.FQ-34	LTM.FQ-33	LTM.FQ-32	LTM.FQ-31	...	LTM.FQ-9	LTM.FQ-8	LTM.FQ-7	LTM.FQ-6	LTM.FQ-5	LT
27	Revenue (Reported)	2490381	2505242	2515737	2520413	2524884	2423669	2392178	2247713	2103604	...	1705379	1672337	1686699	1733825	1789068	18

1 rows × 41 columns

Image 4 is the code for subsetting only the required columns for further analysis.

Image 5

```
In [81]: #Converting float to integer to remove decimals
         Revenue_final["Revenue_2015"] = Revenue_final["Revenue_2015"].apply(np.int64)
         Revenue_final["Revenue_2016"] = Revenue_final["Revenue_2016"].apply(np.int64)
         Revenue_final["Revenue_2017"] = Revenue_final["Revenue_2017"].apply(np.int64)
         Revenue_final["Revenue_2018"] = Revenue_final["Revenue_2018"].apply(np.int64)
         Revenue_final["Revenue_2019"] = Revenue_final["Revenue_2019"].apply(np.int64)
         Revenue_final["Revenue_2020"] = Revenue_final["Revenue_2020"].apply(np.int64)
         Revenue_final["Revenue_2021"] = Revenue_final["Revenue_2021"].apply(np.int64)

In [82]: #Final Revenue data set
         Revenue_final
```

Out[82]:

	Fastfood_chain	Revenue_2015	Revenue_2016	Revenue_2017	Revenue_2018	Revenue_2019	Revenue_2020	Revenue_2021
0	wendys	1870297	1435418	1223408	1589936	1709002	1733825	1896998
1	starbucks	19162700	21675300	22386800	24719500	26973000	23518000	29060600
2	chipotle	4501223	3904384	4476412	4864985	5586369	5984634	7547061
3	Mcdonalds	25413000	24621900	22820400	21257900	21364400	19207800	23222900
4	dominos	2118295	2394376	2787979	3432867	3618774	3911196	4370727

Image 5 is the code for converting the datatype of the columns and the final data set with the revenue of Wendy's, Starbucks, Chipotle, McDonalds and Domino's.

4. A personal consumption data set is found per state with the data of all the industries, from which we have extracted the food and beverage industry data. After sorting and cleaning the data, following the similar steps as the other datasets.

Image 6

jupyter Wrangling_project_final (1) (autosaved)

In [526]:
```
#Final data set with personal consumption data of all the states in required years
personal_consumption_final.head()
```

Out[526]:

	State Name	personal_consumption_2015	personal_consumption_2016	personal_consumption_2017	personal_consumption_2018	personal_consumption_2019
0	Alabama	12471.7	12741.1	13003.7	13425.9	13851.2
1	Alaska	2773.3	2721.7	2707.8	2741.0	2778.2
2	Arizona	18537.3	19046.5	20009.6	20494.1	20787.4
3	Arkansas	7206.8	7445.5	7595.0	7678.9	7729.2
4	California	113368.8	116078.5	119796.3	123086.1	127768.6

5. To enrich the data we have, for a better analysis we have used the following formula to find the per capita income of people in each state. As we have the data sets for GDP and personal consumption, the formula is GDP/population to get the percapita income value per each state.

Image 7

In [251]:
```
#percapita income is GDP divided by population of USA
population_final['percapita_2015'] = GDP_final['GDP_2015']/population_final['pop_2015']
```

In [252]:
```
population_final['percapita_2016'] = GDP_final['GDP_2016']/population_final['pop_2016']
```

In [253]:
```
population_final['percapita_2017'] = GDP_final['GDP_2017']/population_final['pop_2017']
population_final['percapita_2018'] = GDP_final['GDP_2018']/population_final['pop_2018']
population_final['percapita_2019'] = GDP_final['GDP_2019']/population_final['pop_2019']
population_final['percapita_2020'] = GDP_final['GDP_2020']/population_final['pop_2020']
population_final['percapita_2021'] = GDP_final['GDP_2021']/population_final['pop_2021']
```

In [254]:
```
population_final.head()
```

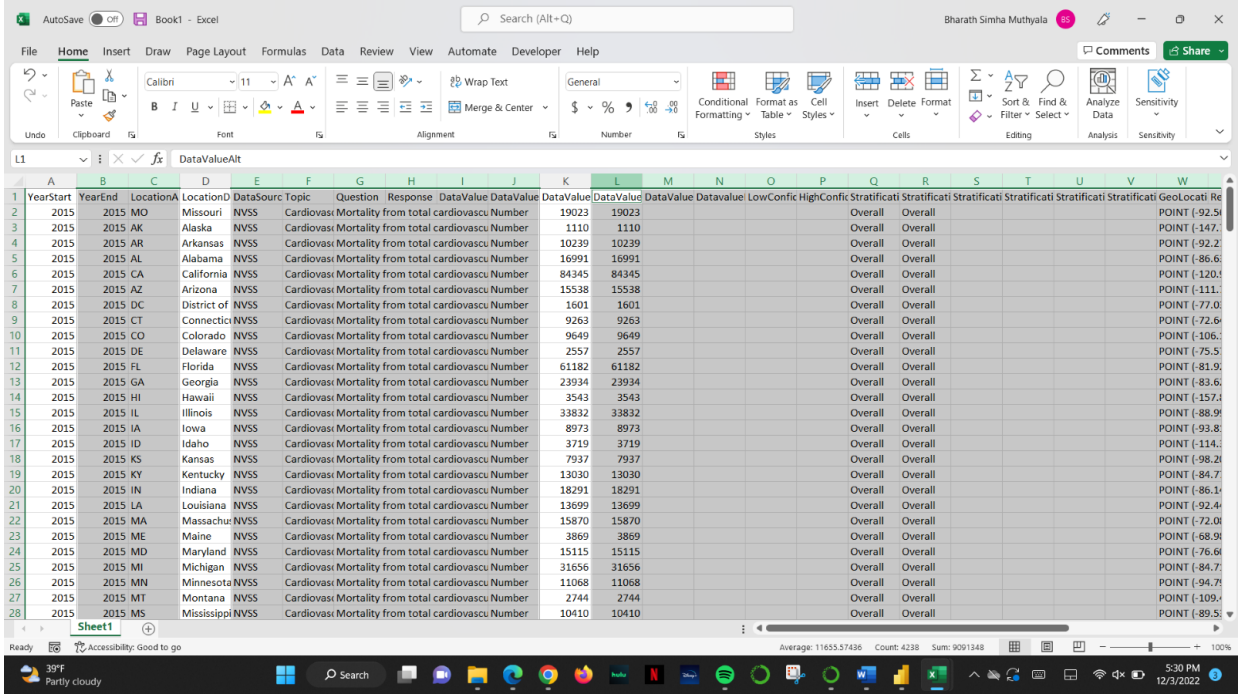
Out[254]:

_2017	pop_2018	pop_2019	pop_2020	pop_2021	percapita_2016	percapita_2017	percapita_2018	percapita_2019	percapita_2020	percapita_2021	percapita_2015
74747	4887871	4903185	5024803	5039877	0.040181	0.040729	0.041182	0.041701	0.040678	0.042269	0.039469
39795	737438	731545	732441	732673	0.072618	0.073209	0.072281	0.073250	0.069725	0.069804	0.073861
16270	7171646	7278717	7177986	7276316	0.042495	0.043929	0.044232	0.045435	0.047114	0.049108	0.041735
04279	3013825	3017804	3012232	3025891	0.038060	0.038257	0.038670	0.039152	0.039735	0.041256	0.037853
36653	39557045	39512223	39499738	39237836	0.062591	0.065756	0.067619	0.070172	0.069527	0.075003	0.060536

Image 7 is the code used to enrich the existing GDP and population data sets.

6. Using Excel, the datasets for cardiovascular diseases and diabetes are filtered. Select only mortality of these health complications and sort the data chronologically. Unnecessary columns are deleted.

Image 8

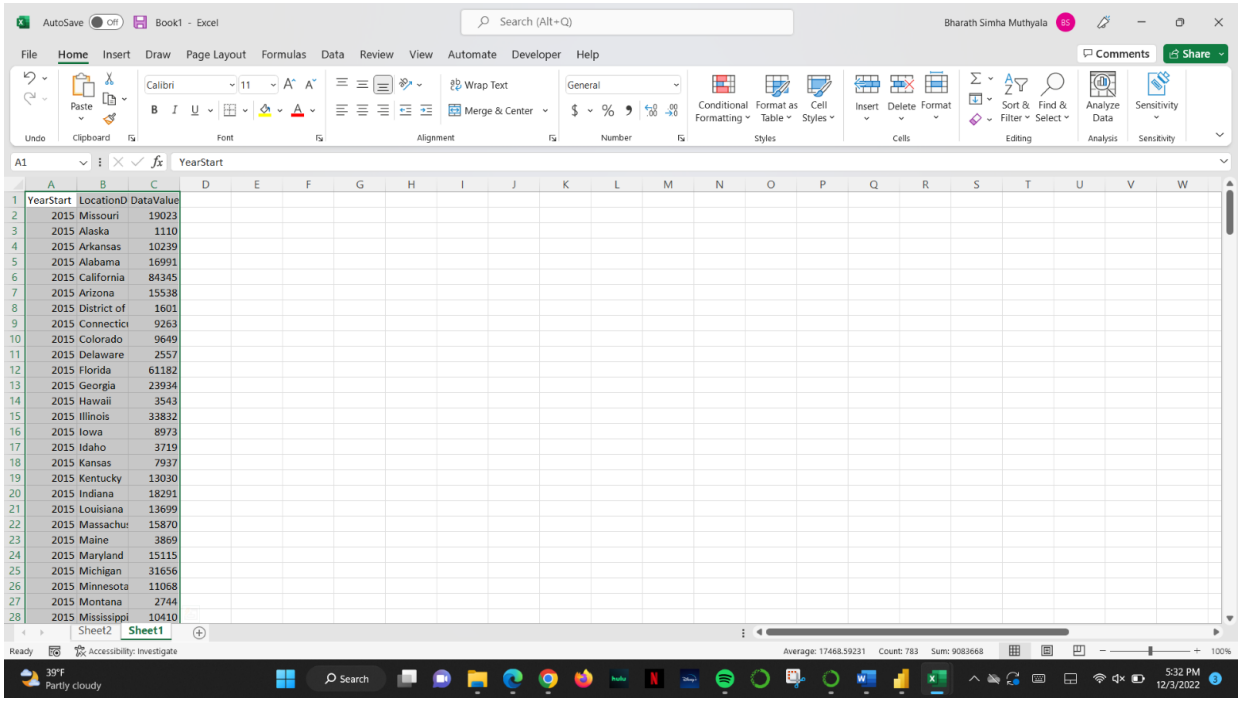


YearStart	YearEnd	LocationA	LocationD	DataSourc	Topic	Question	Response	DataValue	DataValue	DataValue	DataValue	DataValue	Datavalue	LowConfic	HighConfic	Stratificati	Stratificati	Stratificati	Stratificati	Stratificati	Stratificati	GeoLocati	Re
2015	2015	MO	Missouri	NVSS	Cardiovas	Mortality from total cardiovascu		Number		19023	19023					Overall	Overall					POINT (-92.5	
2015	2015	AK	Alaska	NVSS	Cardiovas	Mortality from total cardiovascu		Number		1110	1110					Overall	Overall					POINT (-147.	
2015	2015	AR	Arkansas	NVSS	Cardiovas	Mortality from total cardiovascu		Number		10239	10239					Overall	Overall					POINT (-92.2	
2015	2015	AL	Alabama	NVSS	Cardiovas	Mortality from total cardiovascu		Number		16991	16991					Overall	Overall					POINT (-86.6	
2015	2015	CA	California	NVSS	Cardiovas	Mortality from total cardiovascu		Number		84345	84345					Overall	Overall					POINT (-120.9	
2015	2015	AZ	Arizona	NVSS	Cardiovas	Mortality from total cardiovascu		Number		15538	15538					Overall	Overall					POINT (-111.	
2015	2015	DC	District of	NVSS	Cardiovas	Mortality from total cardiovascu		Number		1601	1601					Overall	Overall					POINT (-77.0	
2015	2015	CT	Connectic	NVSS	Cardiovas	Mortality from total cardiovascu		Number		9263	9263					Overall	Overall					POINT (-72.6	
2015	2015	CO	Colorado	NVSS	Cardiovas	Mortality from total cardiovascu		Number		9649	9649					Overall	Overall					POINT (-106.	
2015	2015	DE	Delaware	NVSS	Cardiovas	Mortality from total cardiovascu		Number		2557	2557					Overall	Overall					POINT (-75.5	
2015	2015	FL	Florida	NVSS	Cardiovas	Mortality from total cardiovascu		Number		61182	61182					Overall	Overall					POINT (-81.9	
2015	2015	GA	Georgia	NVSS	Cardiovas	Mortality from total cardiovascu		Number		23934	23934					Overall	Overall					POINT (-83.6	
2015	2015	HI	Hawaii	NVSS	Cardiovas	Mortality from total cardiovascu		Number		3543	3543					Overall	Overall					POINT (-157.	
2015	2015	IL	Illinois	NVSS	Cardiovas	Mortality from total cardiovascu		Number		33832	33832					Overall	Overall					POINT (-88.9	
2015	2015	IA	Iowa	NVSS	Cardiovas	Mortality from total cardiovascu		Number		8973	8973					Overall	Overall					POINT (-93.8	
2015	2015	ID	Idaho	NVSS	Cardiovas	Mortality from total cardiovascu		Number		3719	3719					Overall	Overall					POINT (-114.	
2015	2015	KS	Kansas	NVSS	Cardiovas	Mortality from total cardiovascu		Number		7937	7937					Overall	Overall					POINT (-98.2	
2015	2015	KY	Kentucky	NVSS	Cardiovas	Mortality from total cardiovascu		Number		13030	13030					Overall	Overall					POINT (-84.7	
2015	2015	IN	Indiana	NVSS	Cardiovas	Mortality from total cardiovascu		Number		18291	18291					Overall	Overall					POINT (-86.1	
2015	2015	LA	Louisiana	NVSS	Cardiovas	Mortality from total cardiovascu		Number		13699	13699					Overall	Overall					POINT (-92.4	
2015	2015	MA	Massachus	NVSS	Cardiovas	Mortality from total cardiovascu		Number		15870	15870					Overall	Overall					POINT (-72.0	
2015	2015	ME	Maine	NVSS	Cardiovas	Mortality from total cardiovascu		Number		3869	3869					Overall	Overall					POINT (-68.9	
2015	2015	MD	Maryland	NVSS	Cardiovas	Mortality from total cardiovascu		Number		15115	15115					Overall	Overall					POINT (-76.6	
2015	2015	MI	Michigan	NVSS	Cardiovas	Mortality from total cardiovascu		Number		31656	31656					Overall	Overall					POINT (-84.7	
2015	2015	MN	Minnesota	NVSS	Cardiovas	Mortality from total cardiovascu		Number		11068	11068					Overall	Overall					POINT (-94.7	
2015	2015	MT	Montana	NVSS	Cardiovas	Mortality from total cardiovascu		Number		2744	2744					Overall	Overall					POINT (-109.	
2015	2015	MS	Mississippi	NVSS	Cardiovas	Mortality from total cardiovascu		Number		10410	10410					Overall	Overall					POINT (-89.5	

Image 8 is the process we followed to filter the. Cardiovascular and diabetes mortality data using Excel

The refined data is as follows:

Image 9

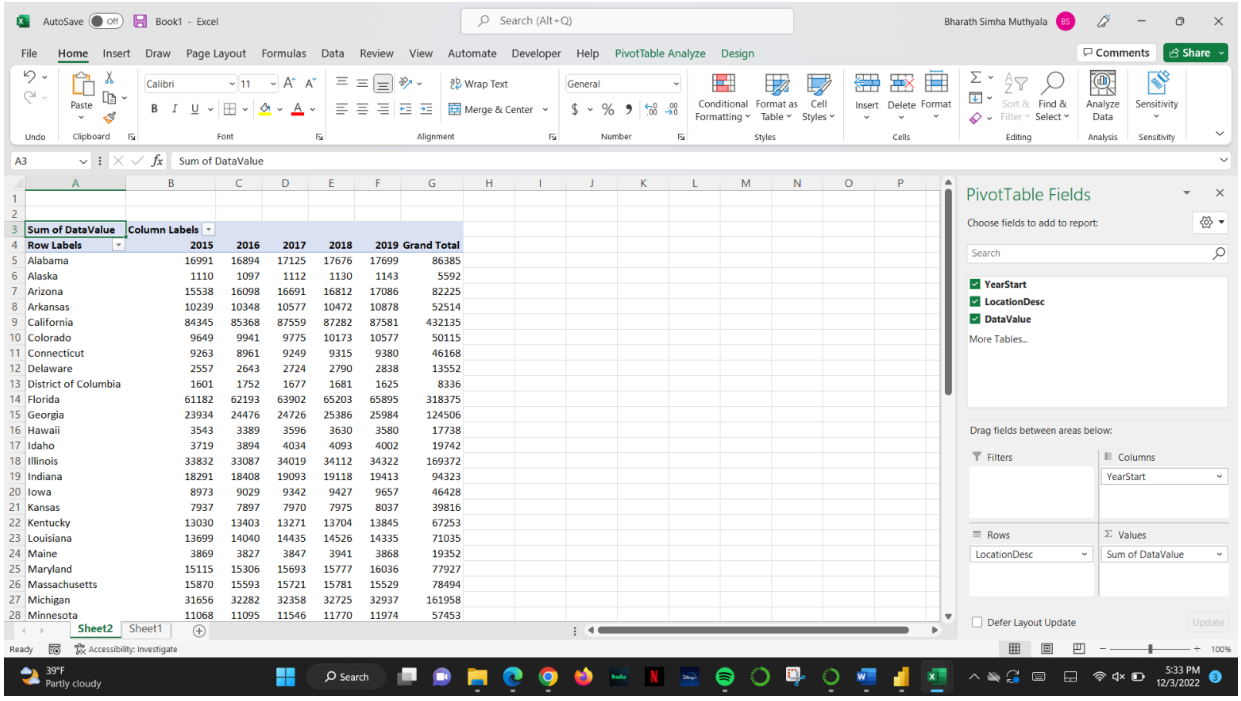


YearStart	LocationD	DataValue
2015	Missouri	19023
2015	Alaska	1110
2015	Arkansas	10239
2015	Alabama	16991
2015	California	84345
2015	Arizona	15538
2015	District of	1601
2015	Connectic	9263
2015	Colorado	9649
2015	Delaware	2557
2015	Florida	61182
2015	Georgia	23934
2015	Hawaii	3543
2015	Illinois	33832
2015	Iowa	8973
2015	Idaho	3719
2015	Kansas	7937
2015	Kentucky	13030
2015	Indiana	18291
2015	Louisiana	13699
2015	Massachus	15870
2015	Maine	3869
2015	Maryland	15115
2015	Michigan	31656
2015	Minnesota	11068
2015	Montana	2744
2015	Mississippi	10410

Image 9 is the final data set obtained after the wrangling process.

7. Using Pivot table, the data was formatted and sorted in a readable and easy way to understand the format.

Image10

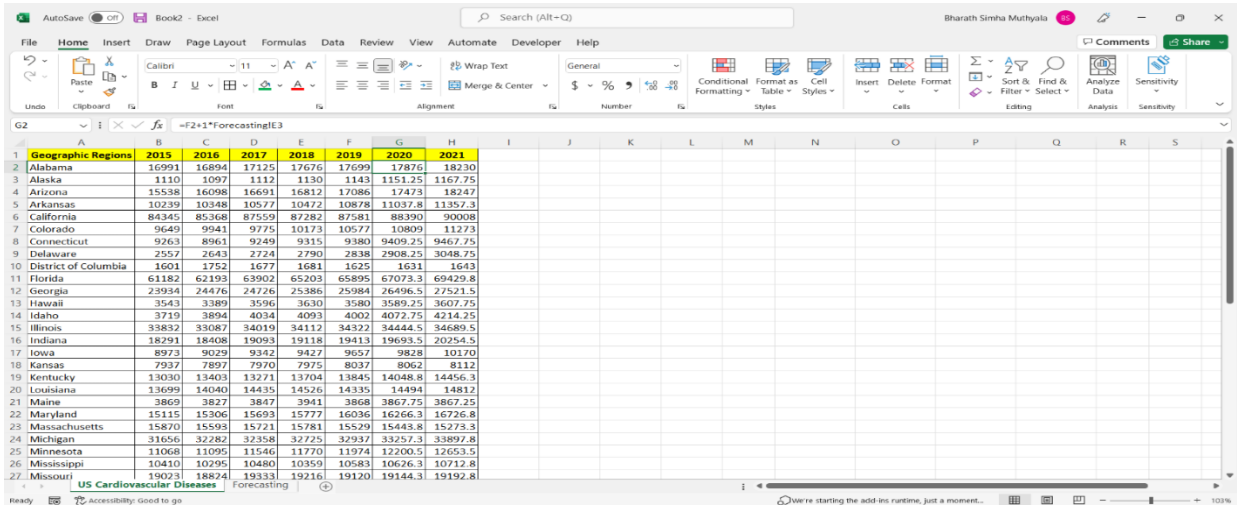


Sum of DataValue	Column Labels					
Row Labels	2015	2016	2017	2018	2019	Grand Total
Alabama	16991	16894	17125	17676	17699	86385
Alaska	1110	1097	1112	1130	1143	5592
Arizona	15538	16098	16691	16812	17086	82225
Arkansas	10239	10348	10577	10472	10878	52514
California	84345	85368	87559	87282	87581	432135
Colorado	9649	9941	9775	10173	10577	50115
Connecticut	9263	8961	9249	9315	9380	46168
Delaware	2557	2643	2724	2790	2838	13552
District of Columbia	1601	1752	1677	1681	1625	8336
Florida	61182	62193	63902	65203	65895	318375
Georgia	23934	24476	24726	25386	25984	124506
Hawaii	3543	3389	3596	3630	3580	17738
Idaho	3719	3894	4034	4093	4002	19742
Illinois	33832	33087	34019	34112	34322	169372
Indiana	18291	18408	19093	19118	19413	94323
Iowa	8973	9029	9342	9427	9657	46428
Kansas	7937	7897	7970	7975	8037	39816
Kentucky	13030	13403	13271	13704	13845	67253
Louisiana	13699	14040	14435	14526	14335	71035
Maine	3869	3827	3847	3941	3868	19352
Maryland	15115	15306	15693	15777	16036	77927
Massachusetts	15870	15593	15721	15781	15529	78494
Michigan	31656	32282	32358	32725	32937	161958
Minnesota	11068	11095	11546	11770	11974	57453

Image 10 is the process of using pivot table to format the data

8. In the data set, cardiovascular mortality data we found the data for the years as the data set doesn't contain mortality rate of cardiovascular diseases and diabetes for years 2020 and 2021. To enrich the missing data, we have used Forecasting Analysis under Arithmetic Increase Method (Pn=Po+nX) to calculate the final forecasting mortality number for these years and enrich the data.

Image 11



Geographic Regions	2015	2016	2017	2018	2019	2020	2021
Alabama	16991	16894	17125	17676	17699	17876	18230
Alaska	1110	1097	1112	1130	1143	1151.25	1167.75
Arizona	15538	16098	16691	16812	17086	17473	18247
Arkansas	10239	10348	10577	10472	10878	11037.8	11357.3
California	84345	85368	87559	87282	87581	88390	90008
Colorado	9649	9941	9775	10173	10577	10809	11273
Connecticut	9263	8961	9249	9315	9380	9409.25	9467.75
Delaware	2557	2643	2724	2790	2838	2908.25	3048.75
District of Columbia	1601	1752	1677	1681	1625	1631	1643
Florida	61182	62193	63902	65203	65895	67073.3	69429.8
Georgia	23934	24476	24726	25386	25984	26496.5	27521.5
Hawaii	3543	3389	3596	3630	3580	3589.25	3607.75
Idaho	3719	3894	4034	4093	4002	4072.75	4214.25
Illinois	33832	33087	34019	34112	34322	34444.5	34689.5
Indiana	18291	18408	19093	19118	19413	19693.5	20254.5
Iowa	8973	9029	9342	9427	9657	9828	10170
Kansas	7937	7897	7970	7975	8037	8062	8112
Kentucky	13030	13403	13271	13704	13845	14048.8	14456.3
Louisiana	13699	14040	14435	14526	14335	14494	14812
Maine	3869	3827	3847	3941	3868	3867.75	3867.25
Maryland	15115	15306	15693	15777	16036	16266.3	16726.8
Massachusetts	15870	15593	15721	15781	15529	15443.8	15273.3
Michigan	31656	32282	32358	32725	32937	33257.3	33897.8
Minnesota	11068	11095	11546	11770	11974	12200.5	12653.5
Mississippi	10410	10295	10480	10359	10583	10626.3	10712.8
Missouri	19023	18824	19333	19216	19120	19144.3	19192.8

Image 11 is the data enrichment process we followed to forecast the missing data in the cardiovascular mortality table using the existing data.

Analysis & Results:

1. Some of the challenges faced during the wrangling process was the large amount of chronic disease index data from CDC.
2. The data was too large for python to handle as the file size was more than 1 GB.
3. Hence, Excel and power Query was used to pull in the required fields from data and sort/ filter just the data that we require.
4. Once we extracted the relevant fields for Cardiovascular and Diabetes, another challenge faced was that there was no data for 2020 and 2021 and we were unable to source the state-wise data anywhere. Therefore, the forecasting method of Arithmetic Increase was used to calculate the missing years data using historical data from 2015-2019.
5. The year-on-year difference and its percentage from 2015-2019 was calculated. The average difference was known over the years and the Arithmetic Increase formula to fill was used in the missing data.
6. After profiling and analyzing the data, the results suggested that on an average, an American spends $20,000 on off premises food and beverages consumption
7. There is a position correlation between the Population, GDP and per capita income. They all are moving in the same upward direction. In 2020, there has been a slight dip in the GDP and PCI due to the pandemic but there has been a constant increase in the population.
8. There is a positive correlation between the data. As the personal consumption standard of the public is rising over the years, there is an increase in cardiovascular diseases and diabetes' mortality over the years, recently the cardiovascular mortality has been moving rapidly in an upward trend.
9. The revenue of the fast-food restaurants has been increasing in an upward fashion along with cardiovascular disease and diabetes.

External Materials:

In recent years, people's health metrics has been correlated with the increase in fast food joints. We have taken a few external materials and articles supporting our analysis. A few of the articles have been listed down below:

1. https://www.hearthousenj.com/learning-center/diet-nutrition/the-effects-of-fast- food-on-the-heart/ - :~:text=Consuming unhealthy foods at fast,blood pressure and dehydrate you

   This article talks about how consuming unhealthy foods at fast food joints can increase one's chance of obesity and, in turn, increase their risk of diseases associated with excess weight, such as heart disease and diabetes. This article claims that high sodium intake can even increase one's blood pressure.

2. https://www.medicalnewstoday.com/articles/324847#short-term-impacts

   The article proclaims that the processed carbs and added sugar in fast food helps in quick digestion and causes a quick rise in blood sugar. This leads to an unnaturally big spike in insulin, which in turn causes the blood sugar to fall. It may make people feel worn out. Within a short period of time after eating, insulin encourages more hunger.

3. https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2008.137638

   This article examined the relationship between fast-food restaurants near schools and obesity among middle and high school students in California. For the results, information from individual-level student responses to the 2002–2005 California Healthy Kids Survey (CHKS) was used. The primary outcome of interest was BMI. They also considered binary outcomes for overweight and obesity. The obesity measurements of those younger than 19 years were based on percentiles by age and gender reference group, according to the BMI-for-age percentiles chart published by the Centre for Disease Control and Prevention (CDC). A child at or above the 85th percentile of BMI distribution by age and gender was considered overweight. A child at or above the 95th percentile was considered obese (and overweight).

4. https://www.diabetes.co.uk/news/2006/jan/fast-food-and-diabetes-link.html

   This research paper on "*Diabetes and Fast-Food Link*" is read to know and understand more about the positive relation between these two variables.

5. https://www.zippia.com/advice/us-fast-food-industry-statistics/

   The article helped us understand more about the facts and figures of the United States' fast-food trends and statistics which claims a steep rise in its market size and revenue.

6. https://www.medicalnewstoday.com/articles/317122 - effects

Junk foods are high in trans and saturated fats, which can raise levels of triglycerides, a type of fat that ispresent in the blood. High levels of triglycerides increase the risk of developing type 2 diabetes.

7. https://www.ibisworld.com/united-states/market-research-reports/fast-food-restaurants-industry/

The study claims that over the next five years to 2027, the fast-food restaurant sector is projected to continue to have a significant impact on the food service industry. As consumers continue to look for quick and inexpensive meal options, the industry's capacity to offer convenient food at a low price is likely to stay popular in the years to come.

8. https://www.sciencedaily.com/releases/2020/02/200214134723.htm

According to a new Dartmouth-led study, there is a strong link between the amount of fast food that pre-school age children consumeand their likelihood of becoming overweight or obese.

9. https://www.statista.com/statistics/1174417/fast-food-restaurants-industry-market-size-us/

Numerous fast-food chains, including McDonald's, are well-known names both abroad and in the United States. McDonald's topped a ranking of American fast-food restaurants by sales in 2021 by more than 20 billion dollars, and it was followed by Starbucks, Chick-fil-A, Taco Bell, and Subway. McDonald's came in third behind Starbucks and Subway was by far the QSR with the most units in the United States in 2021.

10. https://libcom.org/history/economics-fast-food-industry

The article suggests that the consumers are eating at fast food restaurants more frequently than at other eateries, which is why the burger franchises are doing well.

11. https://www.medicalnewstoday.com/articles/324847#short-term-impacts

Due to the processed carbs and added sugar in fast food, it digests quickly and causes a quick rise in blood sugar. This leads to an unnaturally big spike in insulin, which in turn causes the blood sugar to fall. It may make people feel worn out. Within a short period of time after eating, insulin encourages more hunger.

Future Potential Data and Analysis:

- For an additional analysis, the state-wise revenue data of fast-food restaurants would have been collected. The data wrangling techniques are applied only on the top five United States' quick serving restaurants.

- The project work included analysis of only two health complications, cardiovascular diseases and diabetes. If we had more time, he work would have included in-depth analysis to correlate more diseases with the fast-food consumption.