**Save the Tatas – Breast Cancer Capstone**

**Contents**

* Motivation
* Data Sources and Tools
* Challenges
* Conclusions

**Motivation**

My passion in helping others has led me to pursue a data analytics career within the healthcare industry. Breast Cancer is in my family and for us, we are fortunate that my cousin was blessed to have access to the medical treatment for her cancer, which is now in remission. Her journey and determination in beating the cancer inspired me to complete this project. I realized that unlike my cousin, there are several females who were not as fortunate to beat their cancer due to several factors, such as the severity of the cancer, financial hurdles, or lack of access to adequate care. The drive for this project is to examine some of those factors and to continue to highlight the importance of combating the inequities of breast cancer deaths amongst our female population.

According to the [State Cancer Profile](https://statecancerprofiles.cancer.gov/quick-profiles/index.php?statename=tennessee#t=3) site, from 2016-2020 Tennessee has a 21.6 mortality rate compared to the USA rate of 19.6. There are several factors that play into why women are dying at higher rates in certain Tennessee counties, one being having access to adequate healthcare facilities and treatment and two being insurance coverage. For this project, I want to examine the mortality rates of women in the state to see if there is a correlation between the high death rates, the number of available mammograph facilities and insurance. For counties that have a higher death rate, how many mammograph facilities are available? For the counties that have a higher death rate, how many of the females within those counties have insurance coverage?

**Data Sources and Tools**

Death Rates by County - [State Cancer Profiles](https://statecancerprofiles.cancer.gov/deathrates/index.php?stateFIPS=47&areatype=county&cancer=055&race=00&sex=2&age=001&type=death#results)

The stats are produced are from a collaboration between the National Cancer Institute and Centers for Disease Control and Prevention.

The data compile looks at death rates by County during the 2016-2020 timeframe.

Insurance Coverage & Income – [Census Bureau](https://data.census.gov/)

US Census Bureau – American Community Survey 5-Year Estimate 2016 – 2020

FDA Mammograph Facilities – [FDA U.S. Food & Drug Administration](https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMQSA/mqsa.cfm)

**Tools**

Excel & Python/Jupyter Notebooks for data cleaning and analysis

Tableau for data visualizations.

**Data Analysis**

For my data analysis, I extracted the death rates in an Excel file, which contained calculated death rates for each county, from the State Cancer Profiles.

The census data was obtained from an API to have a more machine-readable dataset. The API and instructions can be found [here](https://www.census.gov/data/developers/data-sets.html). The dataset had to be converted from a JSON file to CSV file to make few changes to the data, such as using vlookup to match the correct column headers to its respective variable labels. Once the data was gathered, I used Python/Jupyter Notebook to create a subset dataframe with only the columns needed for analysis. Before merging the census data with the death rates dataframe, I had to reduce the row counts within the census data by creating columns that contain a total sum of insured and uninsured females and household totals by income from 2016-2020 for each county. Once those calculated columns were created, the new dataset was merged with the death rate on county name. The new dataset now contains the death rates by county, the total number of insured and uninsured females and household total by income. This new dataset was used to create my visualizations.

The mammograph facilities was extracted from the FDA website as a PDF and later converted into an Excel file. It would a later be combined with the death rate and census data in Tableau for visualization.

**Challenges**

The main challenge I faced was the lack of female death rates by race for all TN counties due to data suppression. Tennessee has 95 counties and data was only available for 69. The data has been suppressed to ensure confidentiality and stability of rate estimates. Counts are suppressed if fewer than 16 records were reported in a specific area-sex-race category[1](https://statecancerprofiles.cancer.gov/deathrates/index.php?stateFIPS=47&areatype=county&cancer=055&race=00&sex=2&age=001&type=death#results).

Due to the data suppression, it was quite difficulty to get a more defined looked at the death rates by races. Initially, I wanted to look and compare mortality rates amongst white, black, and brown women to highlight some of the healthcare disparities within our community by race, but due to how the data was suppressed, I couldn’t paint that full picture for this project.

So, I had to switch gears and change my focus to look at mortality rates amongst all women in Tennessee.

Also, due to time constraints, I couldn’t provide an additional breakdown for household by income by gender. However, that data will be provide at a later time.

**Conclusion**

For the PowerPoint presentation, I limited the counties count to 12, but for the ReadMe, I wanted to highlight the top 15.

**Which counties had the highest BC death rates?**

1. Crockett County - 31.9

2. Unicoi County - 31.7

3. Scott County - 31.3

4. Weakley County - 31.1

5. Grundy County - 30.9

6. Henderson County - 29.7

7. Tipton County - 29.6

8. Lauderdale County - 29.4

9. Union County - 27.3

10. Dickson County - 27.2

11. Shelby County - 27

12. Hickman County - 26.7

13. Chester County - 26.6

14. Morgan County - 26.5

15. Cheatham County - 25.9

In looking at these counties with the highest death rate, I wanted to see how many FDA Mammograph facilities are located within these counties?

In TN, there are 192 FDA approved Mammograph facilities available within 69 counties. Within the 15 counties with the highest mortality rates, only 8 of the counties have a mammograph facility. However, I did notice an outlier with Shelby County. Within that county, there are 26 mammograph facilities, the largest number within the entire state, but has a high mortality rate of 27%.

Jumping to my first data question, is there a correlation between the counties that have a higher death rate with the number of available mammograph facilities? Based on the available data, the answer is yes. Within the 15 counties with the highest death rate, only 8 of the 15 have 1 or fewer than 2 mammograph facilities, apart from Shelby County. Which lead me to think, “what is going on within Shelby County”?

This brings me to my second data question, is there another factor, such as insurance that could also plays a factor in higher mortality rates within these counties? The answer is yes.

According to the US Census Data, between 2016-2020, Shelby County has the highest number of its population with no insurance coverage at 4,052,700 and out of that number 264,365 are females. The highest number of uninsured females within the State during the that timeframe.

**Which counties had the lowest BC death rates?**

1. Roane County -14.6

2. Hardin County - 15

3. Greene County - 15

4. Carter County - 16

5. Franklin County - 16.9

6. Fayette County - 16.9

7. Maury County - 17.4

8. Robertson County - 17.7

9. Henry County - 18.6

10. Jefferson County -18.7

11. Rhea County - 19

12. Blount County - 19.1

13. Knox County - 19.3

14. Montgomery County - 19.4

15. Cocke County - 19.4

In looking at these counties with the lowest morality rates, 14 out of the 15 counties had 1 or 2 mammograph facilities, with Knox County leading the pack with 17 facilities.

In looking at the insurance factor, the top 15 had the lowest number of uninsured females during 2016-2020. Knox county, which has the highest number of mammograph facilities in the group, also has the least amount of uninsured females of approx. 91,969, compared to Shelby county’s 264,365. The other counties on the list also had more **insured females** compared to the counties with the highest mortality rate.

**Final Analysis:** Insurance and access to adequate healthcare plays a key part in contributing to high mortality rates of breast cancer amongst women in Tennessee.

I created a dashboard to provide a visual picture of my analysis. I hope from the information provide will help play a small part in highlighting inequity within our healthcare system and encourage our Healthcare administers to continue championing for initiatives to help reduce the Breast Cancer mortality rate to save more female lives within our State.

<https://public.tableau.com/shared/PR5HRCDCB?:display_count=n&:origin=viz_share_link>