|  |
| --- |
|  |
| Impact of Game of Thrones Character Names on US baby names |
|  |

|  |
| --- |
| shrikar Umarji  4/1/2016 |

Contents

[1. Goal: 1](#_Toc449720127)

[2. Data: 1](#_Toc449720128)

[3. SAS Analysis 2](#_Toc449720129)

[3.1. Data Import and cleaning 2](#_Toc449720130)

[3.2. Analysis 4](#_Toc449720131)

[3.3. Individual Analysis 5](#_Toc449720132)

[3.3.1. Arya 5](#_Toc449720133)

[3.3.2. Khaleesi: 6](#_Toc449720134)

[3.3.3. Theon 7](#_Toc449720135)

[3.3.4. Tyrion 8](#_Toc449720136)

[3.3.5. Daenerys 9](#_Toc449720137)

[4. Conclusions 10](#_Toc449720138)

[5. Future Analysis 10](#_Toc449720139)

# 1. Goal:

To analyse the effect on Game of Thrones character Names on the Baby names across United States of America.

# 2. Data:

The Dataset US Baby Names was taken from kaggle.com (https://www.kaggle.com/kaggle/us-baby-names). There are two datasets. One is the NationalNames dataset shows the distribution of names across US while the other data set is StateNames which shows state wise distribution of Names in US.

For this analysis we will be considering NationalNames dataset to see the overall effect on Game of Thrones.

The NationalNames data set contains count of all baby names in US from 1910 to 2014. I will be working on a subset of this data set. Since Game of thrones series first season was aired in 2011, the subset contains baby names from 2008 to 2014.

A R script was written to get the subset of the dataset. The R script is shown below

**i*nput <- read.csv("NationalNames.csv")***

***subset\_data <- subset(input, Year > "2007")***

***write.csv(subset\_data, "Names\_2008.csv")***

This script takes the NationalNames.csv dataset as an input. Creates a subset of the baby names from 2008 to 2014 and writes the subset as Names\_2008.csv.

This is the file that will be imported in SAS for further analysis in SAS.

# 3. SAS Analysis

## 3.1. Data Import and cleaning

I imported the data into SAS using the import procedure in SAS. The code is shown below.

*proc import datafile = "S:\public\_html\practicum\selected\_states.csv"*

*out = test.input\_data*

*dbms = csv;*

*run;*

Below is how the imported dataset looks.

| **Obs** | **VAR1** | **Id** | **Name** | **Year** | **Gender** | **Count** |
| --- | --- | --- | --- | --- | --- | --- |
| 1 | 1587864 | 1587864 | Emma | 2008 | F | 18799 |
| 2 | 1587865 | 1587865 | Isabella | 2008 | F | 18603 |
| 3 | 1587866 | 1587866 | Emily | 2008 | F | 17421 |
| 4 | 1587867 | 1587867 | Olivia | 2008 | F | 17069 |
| 5 | 1587868 | 1587868 | Ava | 2008 | F | 17027 |
| 6 | 1587869 | 1587869 | Madison | 2008 | F | 17017 |
| 7 | 1587870 | 1587870 | Sophia | 2008 | F | 16073 |
| 8 | 1587871 | 1587871 | Abigail | 2008 | F | 15072 |
| 9 | 1587872 | 1587872 | Elizabeth | 2008 | F | 11986 |
| 10 | 1587873 | 1587873 | Chloe | 2008 | F | 11816 |
| 11 | 1587874 | 1587874 | Samantha | 2008 | F | 11175 |
| 12 | 1587875 | 1587875 | Addison | 2008 | F | 10755 |
| 13 | 1587876 | 1587876 | Natalie | 2008 | F | 10193 |
| 14 | 1587877 | 1587877 | Mia | 2008 | F | 10166 |
| 15 | 1587878 | 1587878 | Alexis | 2008 | F | 9713 |
| 16 | 1587879 | 1587879 | Alyssa | 2008 | F | 9626 |
| 17 | 1587880 | 1587880 | Hannah | 2008 | F | 9554 |
| 18 | 1587881 | 1587881 | Ashley | 2008 | F | 9400 |
| 19 | 1587882 | 1587882 | Ella | 2008 | F | 9344 |
| 20 | 1587883 | 1587883 | Grace | 2008 | F | 9032 |

This dataset has a lot of redundant variables which are not going to be used in this analysis. So, we need to clean this data by removing these unnecessary variables.

We will use the DATA step in SAS to perform data cleaning. Below is the code used to clean the dataset.

***data*** *test.cleaned\_data(drop=VAR1 X ID);*

*set test.input\_data;*

***run****;*

This step will created a new dataset cleaned\_data that will not contain the redundant variables VAR1, X and ID.

| **Obs** | **Name** | **Year** | **Gender** | **Count** |
| --- | --- | --- | --- | --- |
| 1 | Emma | 2008 | F | 18799 |
| 2 | Isabella | 2008 | F | 18603 |
| 3 | Emily | 2008 | F | 17421 |
| 4 | Olivia | 2008 | F | 17069 |
| 5 | Ava | 2008 | F | 17027 |
| 6 | Madison | 2008 | F | 17017 |
| 7 | Sophia | 2008 | F | 16073 |
| 8 | Abigail | 2008 | F | 15072 |
| 9 | Elizabeth | 2008 | F | 11986 |
| 10 | Chloe | 2008 | F | 11816 |
| 11 | Samantha | 2008 | F | 11175 |
| 12 | Addison | 2008 | F | 10755 |
| 13 | Natalie | 2008 | F | 10193 |
| 14 | Mia | 2008 | F | 10166 |
| 15 | Alexis | 2008 | F | 9713 |
| 16 | Alyssa | 2008 | F | 9626 |
| 17 | Hannah | 2008 | F | 9554 |
| 18 | Ashley | 2008 | F | 9400 |
| 19 | Ella | 2008 | F | 9344 |
| 20 | Grace | 2008 | F | 9032 |

In this analysis, since I am only concerned with the impact of Game of Thrones character names, I will again create a subset of cleaned\_data and pick only those names whose effect I want to study. For this analysis, I selected the most popular character names. The names I selected are: Arya, Sansa, Khaleesi, Tyrion, Tywin, Theon, and Daenerys.

The SAS code used to create the above subset is shown below. I again used the DATA step.

*data test.got;*

*set test.cleaned\_data;*

*where Name in ("Arya", "Khaleesi", "Sansa", "Tyrion", "Tywin", "Theon", “Daenerys”);*

*run;*

This code creates a new SAS dataset called “got\_new” which contains only the required subset. A sample of the final dataset is as shown below

| **Obs** | **Name** | **Year** | **Gender** | **Count** |
| --- | --- | --- | --- | --- |
| 1 | Arya | 2008 | F | 205 |
| 2 | Arya | 2008 | M | 114 |
| 3 | Tyrion | 2008 | M | 16 |
| 4 | Arya | 2009 | F | 236 |
| 5 | Arya | 2009 | M | 113 |
| 6 | Tyrion | 2009 | M | 17 |
| 7 | Theon | 2009 | M | 7 |
| 8 | Arya | 2010 | F | 273 |
| 9 | Arya | 2010 | M | 87 |
| 10 | Tyrion | 2010 | M | 11 |
| 11 | Arya | 2011 | F | 387 |
| 12 | Khaleesi | 2011 | F | 28 |
| 13 | Arya | 2011 | M | 111 |
| 14 | Tyrion | 2011 | M | 18 |
| 15 | Arya | 2012 | F | 758 |
| 16 | Khaleesi | 2012 | F | 146 |
| 17 | Daenerys | 2012 | F | 21 |
| 18 | Arya | 2012 | M | 77 |
| 19 | Tyrion | 2012 | M | 20 |
| 20 | Theon | 2012 | M | 12 |

The further analysis is done on this new dataset “got\_new”.

## 3.2. Analysis

I used the REPORT Procedure in SAS to create a tabular summary. The code used is shown below:

***proc******report*** *data=test.got\_new;*

*column Year Name Gender Count;*

*define Name/order;*

*define Year/order;*

***run****;*

The summary obtained is as shown below:

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2008 | Arya | F | 205 |
|  |  | M | 114 |
|  | Tyrion | M | 16 |
| 2009 | Arya | F | 236 |
|  |  | M | 113 |
|  | Theon | M | 7 |
|  | Tyrion | M | 17 |
| 2010 | Arya | F | 273 |
|  |  | M | 87 |
|  | Tyrion | M | 11 |
| 2011 | Arya | F | 387 |
|  |  | M | 111 |
|  | Khaleesi | F | 28 |
|  | Tyrion | M | 18 |
| 2012 | Arya | F | 758 |
|  |  | M | 77 |
|  | Daenerys | F | 21 |
|  | Khaleesi | F | 146 |
|  | Theon | M | 12 |
|  | Tyrion | M | 20 |
| 2013 | Arya | F | 1137 |
|  |  | M | 77 |
|  | Daenerys | F | 68 |
|  | Khaleesi | F | 243 |
|  | Sansa | F | 11 |
|  | Theon | M | 15 |
|  | Tyrion | M | 43 |
| 2014 | Arya | F | 1544 |
|  |  | M | 82 |
|  | Daenerys | F | 86 |
|  | Khaleesi | F | 368 |
|  | Sansa | F | 12 |
|  | Theon | M | 11 |
|  | Tyrion | M | 60 |
|  | Tywin | M | 5 |

There are few very interesting findings.

1. The name Arya appears in all the years and is for both Males and Females.

2. Like Arya, Tyrion also appears during all the years.

3. the most important trend I feel is that the popularity of the name depends on the character’s plotline.

To get more insights, I looked at each of these names individually. Looking at the names individually, a lot of subtle trends came to sight.

## 3.3. Individual Analysis

To perform individual analysis, the SAS REPORT procedure was used. The code used is shown below:

***proc report data=test.got\_new;***

***where Name="Arya";***

***column Year Name Gender Count;***

***define Name/order;***

***define Year/order;***

***run;***

This same code was used for all individual analysis with the name value updated in the WHERE clause.

### 3.3.1. Arya

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2008 | Arya | F | 205 |
|  |  | M | 114 |
| 2009 | Arya | F | 236 |
|  |  | M | 113 |
| 2010 | Arya | F | 273 |
|  |  | M | 87 |
| 2011 | Arya | F | 387 |
|  |  | M | 111 |
| 2012 | Arya | F | 758 |
|  |  | M | 77 |
| 2013 | Arya | F | 1137 |
|  |  | M | 77 |
| 2014 | Arya | F | 1544 |
|  |  | M | 82 |

**Takeaways:**

1. For girls the count of Arya is increasing from 2008 to 2014, whereas for boys the count is decreasing. It is visible that since 2011 when the series was aired, the increase in the name is almost exponential.

2. Throughout the series the character of Arya is evolving and becoming stronger. This correlates to the fact that the popularity of the name is increasing.



The orange line is for Girls while the blue line is for boys. Thus looking at the graph it is clear that 2011 onwards there has been an exponential increase in the popularity of the name Arya.

### 3.3.2. Khaleesi:

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2011 | Khaleesi | F | 28 |
| 2012 | Khaleesi | F | 146 |
| 2013 | Khaleesi | F | 243 |
| 2014 | Khaleesi | F | 368 |

**Takeaways:**

1. Unlike Arya, parents started naming their daughter Khaleesi, once the first season of Game of Thrones aired.

2. Also, in the first season the character was a little subdued and weak. But starting 2nd season she started getting stronger and similarly the popularity of the name also increased.



### 3.3.3. Theon

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2009 | Theon | M | 7 |
| 2012 | Theon | M | 12 |
| 2013 | Theon | M | 15 |
| 2014 | Theon | M | 11 |

**Takeaways:**

1. Unlike the first two names, this name did not become so popular but there is an important trend in his name. In the 3rd season (2013) he betrays the Starks, gets captured and tortured and in the next year we can see that his popularity has decreased.

2. It seems that the fact that how strong or weak the person is has an effect on the popularity of the name.



We can clearly see the drop in popularity in 2014.

### 3.3.4. Tyrion

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2008 | Tyrion | M | 16 |
| 2009 | Tyrion | M | 17 |
| 2010 | Tyrion | M | 11 |
| 2011 | Tyrion | M | 18 |
| 2012 | Tyrion | M | 20 |
| 2013 | Tyrion | M | 43 |
| 2014 | Tyrion | M | 60 |

**Takeaways:**

1. We can see that Tyrion appears consistently throughout 2008 to 2014. But from 2013 there is a big surge in the popularity and that also happens to be the season in which he started developing as a powerful character.



We can clearly see that since 2012, the popularity of the name has increased almost exponentially.

### 3.3.5. Daenerys

| **Year** | **Name** | **Gender** | **Count** |
| --- | --- | --- | --- |
| 2012 | Daenerys | F | 21 |
| 2013 | Daenerys | F | 68 |
| 2014 | Daenerys | F | 86 |

**Takeaways:**

1. Like most of the names, parents started naming their daughter Daenerys once the Game of Thrones series started airing.

2, Daenerys and Khaleesi are names of the same character. But Daenerys could not match Khaleesi in popularity.



The popularity of the name Daenerys has been increasing over time.

# 4. Conclusions

We can draw many different conclusions from this analysis.

1. Game of Thrones definitely has an influence on baby names. Khaleesi and Tyrion are the most popular names which started appearing when Game of Thrones actually started.

2. The popularity of a particular character name depends on how strong or weak the character is in the series. The stronger the character more the popularity of his name.

3. In case of 1 character having different names, some are more popular than others.

# 5. Future Analysis

In this analysis I have just scrapped the surface and a lot of more detailed analysis is possible.

1. One can do a more detailed demographic analysis by looking at the state wise influence of the Game of Thrones Character names.

2. One can do an emotional analysis by looking at what type of parents tend to name their babies based on a TV series.

3. One can also look at why some names are more popular than others.