Dataset: <https://www.kaggle.com/datasets/crowdflower/twitter-user-gender-classification?resource=download>

The Data:

The dataset contains the following fields:

* *unit*id: a unique id for user
* \_golden: whether the user was included in the gold standard for the model; TRUE or FALSE
* *unit*state: state of the observation; one of *finalized* (for contributor-judged) or *golden* (for gold standard observations)
* *trusted*judgments: number of trusted judgments (int); always 3 for non-golden, and what may be a unique id for gold standard observations
* *last*judgment\_at: date and time of last contributor judgment; blank for gold standard observations
* gender: one of *male*, *female*, or *brand* (for non-human profiles)
* gender:confidence: a float representing confidence in the provided gender
* profile\_yn: "no" here seems to mean that the profile was meant to be part of the dataset but was not available when contributors went to judge it
* profile\_yn:confidence: confidence in the existence/non-existence of the profile
* created: date and time when the profile was created
* description: the user's profile description
* fav\_number: number of tweets the user has favorited
* gender\_gold: if the profile is golden, what is the gender?
* link\_color: the link color on the profile, as a hex value
* name: the user's name
* profile*yn*gold: whether the profile y/n value is golden
* profileimage: a link to the profile image
* retweet\_count: number of times the user has retweeted (or possibly, been retweeted)
* sidebar\_color: color of the profile sidebar, as a hex value
* text: text of a random one of the user's tweets
* tweet\_coord: if the user has location turned on, the coordinates as a string with the format "[*latitude*, *longitude*]"
* tweet\_count: number of tweets that the user has posted
* tweet\_created: when the random tweet (in the text column) was created
* tweet\_id: the tweet id of the random tweet
* tweet\_location: location of the tweet; seems to not be particularly normalized
* user\_timezone: the timezone of the user

Steps Done:

1. Download the dataset from Kaggle
2. Backed up the raw file and started working on the copy
3. Imported to Microsoft SQL Server Management Studio for sql queries
4. Analyzed the data by locations and genders and content analysis for two words “war” and “peace”
5. The sql-made views were imported to Tableau
6. Individual charts were made for each piece of info obtained with sql
7. Everything was put together on the tableau dashboard
8. Finalized formatting
9. Backed up the dashboard and saved in tableau public