MSDS 6306: Introduction to Data Science - 405

Unit 05

Due: October 9th, 2017 (Monday, One Hour Before Live Session)

Codebook

Description:

Create a codebook which description of each object that was provided as part of Unit05.

Yob2016.txt – the popular name from 2016, the sex and number of children which has those names. Input file

First name: popular name from 2016

Gender: the sex of the child associated with the name

Amount of children: the number of children who was given that name

Yob2015.txt – the popular name from 2015, the sex and number of children which has those names. Input file

First name: popular name from 2015

Gender: the sex of the child associated with the name

Amount of children: the number of children who was given that name

Final: combination of yob2016.txt and yob2015.txt. This file only contains names that can be found in both txt files. Output file

First name: popular name in both 2015 and 2016

Gender: the sex of the children associated with the name in 2016

2015 Total : total found in the 2015 file

Gender: gender of the name in file 2016

2016 Total: total found in the 2016 file

Total: total of 2015 Total and 2016 Total

MyDate.csv: Final csv file containing only the top 10 most popular girl’s.

Name: most popular girl’s name from both 2015 and 2016

Total: how many girls have this name