STAT 19000 Project 9

Topics: The awk utility

Motivation: The awk utility gives us power, directly in the terminal, to do some data analysis. Indeed, awk commands can even be included in pipelines with other UNIX commands.

Context: All awk programs are just one line of code. We can do simple things very easily in awk, for instance, cutting out certain columns, simply arithmetic, dealing with missing data, etc. In this project, we will only introduce the basics of awk, but there are entire books devotes to this little tool.

Scope: Awk is best used with data that is formatted in the same way as a spreadsheet, i.e., with the same number of fields per line, and consistent types of data in each. Awk is also a fast tool that can process data as part of the large data-driven workflow.

Question 1: Analyzing yellow taxi cab rides

Before you start question 1, please consider the examples given here:

`/class/datamine/data/examples/project9examples.txt`

Use this template to submit your solutions:

`/class/datamine/data/examples/project9tempate.txt`

1a. Use awk to solve Project 3, Question 2b, namely:

How many passengers (altogether) rode in yellow taxi cab rides in New York City in June 2019?

1b. Use awk to solve Project 4, Question 2b, namely:

What is the mean total number of passengers in a New York City yellow taxi cab ride in June 2019?

Question 2: Analyzing grocery store transaction data

This question is related to (but not exactly the same as Project 6, Question 3)

2a. Use awk to analyze the 8451 transactions data: find the total amount (in dollars) spent on grocery purchases (altogether) on 23 December 2017.

2b. Use awk to find the average amount (in dollars) spent in a transaction, on 23 December 2017.

Question 3: Analyzing election campaign data

3a. Use awk to find the average amount (in dollars) in a donation in the 2018 election campaign.

Project Submission:

Submit your solutions for the project at this URL: https://classroom.github.com/a/j8RavqUb using the instructions found in the GitHub Classroom instructions folder on Blackboard.