

Data Methodology:

Finding the Most Popular Email Provider During a 5-Day Work Week

By: Kailani-Brionna Bailey

Client: Needing to find out which email provider is used the most during the work week (Monday-Friday) by business professionals.

Data Scientist: Can we determine which popular email provider sees the most business during a 5-day work week?

Business Understanding - Collecting data on a group of select email providers. Which are the email account providers that are being assessed? Which day of the week sees the most emails being sent and received? Should we only include the email accounts of people who are working full-time jobs, who don't work on Saturday and Sunday? Are inactive accounts and spam messages being included or excluded?

Analytic Approach - Using data sets from a couple of the most popular email account providers i.e. Google (Gmail) and Yahoo, to support a descriptive analysis. Making sure these accounts are active and are still in use by analyzing data and determining when their most recent email was sent and/or received.

Data Requirements - Data sets must only include information from active user accounts and information from account owners who are working a full-time job, excluding the weekends (Saturday and Sunday).

Data Collection - Select a time frame (month and year) to collect data on. Receiving the necessary account information and data set files from knowledgeable stakeholders. Taking note of which day of the week sees the most emails being sent and/or received by full-time employees.

Data Understanding and Preparation - Communicating with stakeholders and confirming that the data collected is only from account owners who work a full-time job (excluding weekends) and who have recently accessed their account to send and/or receive work-related emails. Preparing the data by removing, what seems to be, any inactive accounts and any information associated with spam emails and spam (fake email) accounts used for impersonation purposes.

Modeling and Evaluation - Continuously model (iterate) the approved data sets and remove any outlying email account information i.e. inactive accounts or accounts used for malicious purposes (spam or phishing), to determine which day of the week sees the most emails being sent/received by full-time business professionals through either one of the selected popular email account providers. Determine which of the two email account providers sees the most use within a 5-day work week.