

Example Problem Set

Econ 145

Overview

You are working for a local retail firm. They have many, many clients and are interested in some basic information. They have asked you to do some basic analysis on their records.

To Receive Credit

- Save the scripting file (i.e. your R program file) as **test.R**. Make sure your capitalization is correct as the autograder is case-sensitive.
- Save the data and title it **fakedata** (e.g. `assign_2_dat <- read_csv("fakedata.csv")`).
- Make sure all changes to the original dataset are done within the R script.
- Save your PERMID at the top of the Rscripts (i.e. `PERMID <- xxxx`)
- Your one page write-up must be submitted in a .pdf to receive credit.

Part 1: Coding Assignment (4 parts)

- 1) How many missing values are there in each column? How many missing values are there in total? Put your answers in a tibble and name it **missing**. The tibble should look like this:

Table 1: Example Subset of Tibble for Question 1

Class	Num
Names	#
State	#
Invoice..USD..	#
Paid	#
Total	#

* This is made up data for illustrative purposes.

- 2) How many unique states does the company operate in? Are there any typos in the column? If so, correct the typos. Save the corrected dataframe as `\magenta{data_cleaned}`.
- 3) Find the average invoice and the percent of invoices paid. Save them as **avg_inv** and **avg_inv_paid**.
- 4) Create a new tibble where **Names** is split into two columns: **First Name** and **Last Name**. Nothing else needs to be changed. Save this as `.`. Your tibble should look like so:

Table 2: Example Subset of Tibble for Question 3

Last Name	First Name	State	Invoice..USD..	Paid
Degenhart	Misbaah	California	111.151	Yes
al-Hatem	Joreden	Texas	108.892	No
Thomas	Manuel	Utah	94.613	No
Zhen	Chaise	Nevada	102.911	No
Mccurdy	James	Colorado	101.382	Yes

* This is made up data for illustrative purposes.