Chipotle Case Study

Shu Liu

2023-06-02

## Chiptole Sales

This is Data in Motion data analysis challenge #1 More details click here [link](https://d-i-motion.com/lessons/challenge-1-chipotle-sales/)

### Scenario

You are a financial data analyst at Chipotle and your manager has tasked you with analyzing the most recent sales numbers. She has provided the following set of questions she would like answered.

### Get the data

Link to dataset: Link to [dataset](https://raw.githubusercontent.com/justmarkham/DAT8/master/data/chipotle.tsv)

### Challenge Questions

1. Which was the most-ordered item?
2. For the most-ordered item, how many items were ordered?
3. What was the most ordered item in the choice\_description column?
4. How many items were ordered in total?
5. Turn the item price into a float
6. How much was the revenue for the period in the dataset?
7. How many orders were made in the period?
8. What is the average revenue amount per order?
9. How many different items are sold?

## Steps

### Set up environments

Notes: install package “tidyverse”

install.packages("tidyverse", repos = "http://cran.us.r-project.org") #an argument is added to the function that gives it the web address of the repository. Once the data file is downloaded into the proper directory you will then be able to access your newly installed package.

## Installing package into 'C:/Users/liuch/AppData/Local/R/win-library/4.3'  
## (as 'lib' is unspecified)

## package 'tidyverse' successfully unpacked and MD5 sums checked  
##   
## The downloaded binary packages are in  
## C:\Users\liuch\AppData\Local\Temp\RtmpCMGCf0\downloaded\_packages

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)  
library(dplyr) #for sorting

### Load data

Save the dataset into local file directory. change working directory to where the file is. load the data into dataframe chipotle\_sales.

setwd("C:/Users/liuch/OneDrive/文档/DataAnalytics/Portfolio/case\_study\_1")  
chipotle\_sales <- read\_tsv("chipotle.tsv")

## Rows: 4622 Columns: 5  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: "\t"  
## chr (3): item\_name, choice\_description, item\_price  
## dbl (2): order\_id, quantity  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Now the chiptole\_sales has the data. Let’s take a glimpse.

glimpse(chipotle\_sales)

## Rows: 4,622  
## Columns: 5  
## $ order\_id <dbl> 1, 1, 1, 1, 2, 3, 3, 4, 4, 5, 5, 6, 6, 7, 7, 8, 8, …  
## $ quantity <dbl> 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …  
## $ item\_name <chr> "Chips and Fresh Tomato Salsa", "Izze", "Nantucket …  
## $ choice\_description <chr> "NULL", "[Clementine]", "[Apple]", "NULL", "[Tomati…  
## $ item\_price <chr> "$2.39", "$3.39", "$3.39", "$2.39", "$16.98", "$10.…

### Analyze data and answer questions

1. Which was the most-ordered item? 2. For the most-ordered item, how many items were ordered?

chipotle\_sales\_sum <- chipotle\_sales %>%   
 group\_by(item\_name) %>%   
 summarise(ordered\_num=sum(quantity)) ##created a new dataframe for each item and its ordered sum.   
chipotle\_sales\_sum %>% arrange(desc(ordered\_num)) ##use arrange() and desc() to sort desc.

## # A tibble: 50 × 2  
## item\_name ordered\_num  
## <chr> <dbl>  
## 1 Chicken Bowl 761  
## 2 Chicken Burrito 591  
## 3 Chips and Guacamole 506  
## 4 Steak Burrito 386  
## 5 Canned Soft Drink 351  
## 6 Chips 230  
## 7 Steak Bowl 221  
## 8 Bottled Water 211  
## 9 Chips and Fresh Tomato Salsa 130  
## 10 Canned Soda 126  
## # ℹ 40 more rows

**Chicken Bowl** was the most-ordered item. **761** were ordered.

1. What was the most ordered item in the choice\_description column?

chipotle\_sales\_choice <- chipotle\_sales %>% group\_by(choice\_description) %>% summarise(ordered\_num=sum(quantity)) ##create a new dataframe chipotle\_sales\_choice for calculating total ordered number for each different choice.  
chipotle\_sales\_choice %>% filter(choice\_description!="NULL") %>% arrange(desc(ordered\_num))

## # A tibble: 1,043 × 2  
## choice\_description ordered\_num  
## <chr> <dbl>  
## 1 [Diet Coke] 159  
## 2 [Coke] 143  
## 3 [Sprite] 89  
## 4 [Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Let… 49  
## 5 [Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream]] 42  
## 6 [Fresh Tomato Salsa, [Rice, Black Beans, Cheese, Sour Cream, Gua… 40  
## 7 [Fresh Tomato Salsa (Mild), [Pinto Beans, Rice, Cheese, Sour Cre… 36  
## 8 [Lemonade] 36  
## 9 [Coca Cola] 32  
## 10 [Fresh Tomato Salsa, [Rice, Black Beans, Cheese]] 30  
## # ℹ 1,033 more rows

**Diet Coke** is the most ordered item in the choice\_description column

4.How many items were ordered in total?

summarise(chipotle\_sales, item\_sold=sum(quantity))

## # A tibble: 1 × 1  
## item\_sold  
## <dbl>  
## 1 4972

**4972** items were ordered in total.

5.Turn the item price into a float First duplicate the item\_price column to item\_price\_db in case make any mistake.

chipotle\_sales <- mutate(chipotle\_sales, item\_price\_db=item\_price)

Then covert the item\_price\_db to float.

chipotle\_sales$item\_price\_db <- as.numeric(gsub("[^0-9.]","",chipotle\_sales$item\_price\_db))

Now we can see item\_price\_db is float data type.

6.How much was the revenue for the period in the dataset? To do this we need to know multiply item\_price\_db and quantity then get revenue.

chipotle\_sales <- mutate(chipotle\_sales, total = quantity \* item\_price\_db) ##add a new column total for total price sold for each order  
revenue <- chipotle\_sales %>% summarise(sum(total)) ##calculate the total by summarise(sum(total))

The revenue for the period is **$39,237**

7.How many orders were made in the period?

total\_orders <-chipotle\_sales %>% summarise(n\_distinct(order\_id)) ##n\_distinct() is count distinct value

**1834** orders were made in the period

8.What is the average revenue amount per order? To calculate the average revenue amount per order = revenue/total\_orders.

revenue/total\_orders

## sum(total)  
## 1 21.39423

So the average revenue per order is $21.39.

9.How many different items are sold?

summarise(chipotle\_sales,items\_kind=n\_distinct(item\_name))

## # A tibble: 1 × 1  
## items\_kind  
## <int>  
## 1 50