ecommerce_purchase_solutions in R

Sripad Kulkarni

24/11/2020

load the libraries

```
pacman::p_load(tidyverse, stringr)
```

load the dataset

```
ecom <- read_csv("/Users/sripadk/Desktop/Ecommerce")
head(ecom)
```

```
## # A tibble: 6 x 14
##   Address Lot   `AM or PM` `Browser Info` Company `Credit Card` `CC Exp Date`
##   <chr>   <chr> <chr>      <chr>          <chr>           <dbl> <chr>
## 1 "16629~ 46 in PM         Opera/9.56.(X~ Martin~       6.01e15 02/20
## 2 "9374 ~ 28 rn PM         Opera/8.93.(W~ Fletch~       3.34e15 11/18
## 3 "Unit ~ 94 vE PM         Mozilla/5.0 (~ Simpso~       6.76e11 08/19
## 4 "7780 ~ 36 vm PM         Mozilla/5.0 (~ Willia~       6.01e15 02/24
## 5 "23012~ 20 IE AM         Opera/9.58.(X~ Brown,~       6.01e15 10/25
## 6 "7502 ~ 21 XT PM         Mozilla/5.0 (~ Silva--       3.02e13 07/25
## # ... with 7 more variables: `CC Security Code` <dbl>, `CC Provider` <chr>,
## #   Email <chr>, Job <chr>, `IP Address` <chr>, Language <chr>, `Purchase
## #   Price` <dbl>
```

What is the average Purchase Price?

```
mean(ecom$`Purchase Price`, na.rm = T)
```

```
## [1] 50.3473
```

What were the highest and lowest purchase prices? **

```
max(ecom$`Purchase Price`, na.rm = T)
```

```
## [1] 99.99
```

```
min(ecom$`Purchase Price`, na.rm = T)
```

```
## [1] 0
```

How many people have English 'en' as their Language of choice on the website?

```
ecom %>%
  filter(Language == 'en') %>%
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1  1098
```

How many people have the job title of "Lawyer" ? **

```
ecom %>%
  filter(Job == 'Lawyer') %>%
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1    30
```

How many people made the purchase during the AM and how many people made the purchase during PM ?

```
ecom %>%
  count(`AM or PM`) %>%
  ungroup()
```

```
## # A tibble: 2 x 2
##   `AM or PM`     n
##   <chr>      <int>
## 1 AM          4932
## 2 PM          5068
```

What are the 5 most common Job Titles?

```
ecom %>%
  count(Job, sort = T) %>%
  head(5)
```

```
## # A tibble: 5 x 2
##   Job                              n
##   <chr>                        <int>
## 1 Interior and spatial designer    31
## 2 Lawyer                           30
## 3 Social researcher                28
## 4 Designer, jewellery              27
## 5 Purchasing manager               27
```

Someone made a purchase that came from Lot: "90 WT" , what was the Purchase Price for this transaction?

```
ecom %>%
  filter(Lot == '90 WT') %>%
  select(`Purchase Price`)
```

```
## # A tibble: 1 x 1
##   `Purchase Price`
##              <dbl>
## 1             75.1
```

What is the email of the person with the following Credit Card Number: 4926535242672853

```
ecom %>%
  filter(`Credit Card` == 4926535242672853) %>%
  select(`Email`)
```

```
## # A tibble: 1 x 1
##   Email
##   <chr>
## 1 bondellen@williams-garza.com
```

How many people have American Express as their Credit Card Provider *and* made a purchase above $95 ?

```
ecom %>%
  filter(`CC Provider` == 'American Express' & `Purchase Price` > 95) %>%
  count()
```

```
## # A tibble: 1 x 1
##       n
##   <int>
## 1    39
```

How many people have a credit card that expires in 2025?

```r
sum(str_count(ecom$`CC Exp Date`, '25' ))
```

```
## [1] 1033
```

Hard: What are the top 5 most popular email providers/hosts

```r
ecom %>%
  mutate(domain = sub('.*@', "",ecom$Email)) %>%
  count(domain, sort = T) %>%
  head(5)
```

```
## # A tibble: 5 x 2
##   domain           n
##   <chr>        <int>
## 1 hotmail.com   1638
## 2 yahoo.com     1616
## 3 gmail.com     1605
## 4 smith.com       42
## 5 williams.com    37
```