

Jeep Analysis R

Abe

2022-05-07

Jeep Renegade Analysis Summary

This R Markdown Document is about Jeep customer data. This data set was given to me as part of a take home assignment for a job interview. I conducted the initial analysis in Excel and built the visualizations in Tableau.

Loading Dataset & Libraries

Libraries include:

- dplyr
- ggplot2
- readxl

Checking data frame structure

Understanding what variables make up the data frame is crucial prior to analysis. Since, I will be using ggplot later in my analysis I must see what variables need to be factors. Factors are data types used to identify categories. ggplot works best with factors and it is best practice to change any variable to factors prior to analysis.

```
str(jeepdata)
```

```
## tibble [23,900 x 9] (S3: tbl_df/tbl/data.frame)
## $ UNIQUE_VISITORS: num [1:23900] 8 4 4 4 12 2 12 2 2 50 ...
## $ FORM_STARTS    : num [1:23900] 10 4 6 0 2 2 0 2 2 36 ...
## $ FORM_COMPLETES : num [1:23900] 2 0 2 0 6 0 0 0 0 16 ...
## $ CHANNEL        : chr [1:23900] "Auto Sites" "Organic Search" "PPC" "Organic Search" ...
## $ FORM_NAME       : chr [1:23900] "FORM_A" "FORM_B" "FORM_B" "NONE" ...
## $ NEW DATE        : POSIXct[1:23900], format: "2019-04-07" "2019-04-07" ...
## $ BROWSER         : chr [1:23900] "Chrome" "Other" "Safari" "Internet Explorer" ...
## $ COUNTRY         : chr [1:23900] "USA" "USA" "USA" "Canada" ...
## $ VISITOR_TYPE    : chr [1:23900] "New" "Returning" "Returning" "New" ...
```

Variables to be changed to factor

1. CHANNEL
2. FORM_NAME
3. BROWSER
4. COUNTRY
5. VISTOR_TYPE

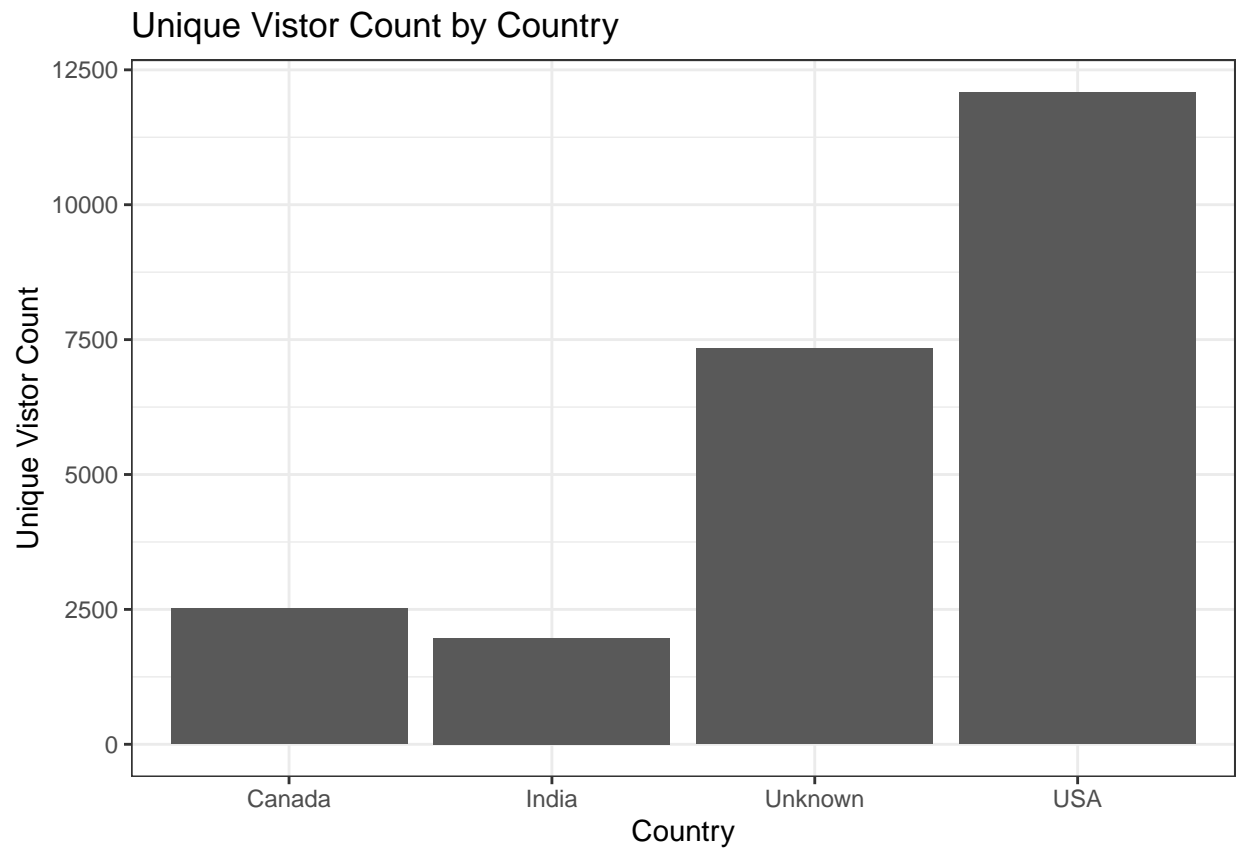
```
str(jeepdata)
```

Verifying the new data types for the above variables

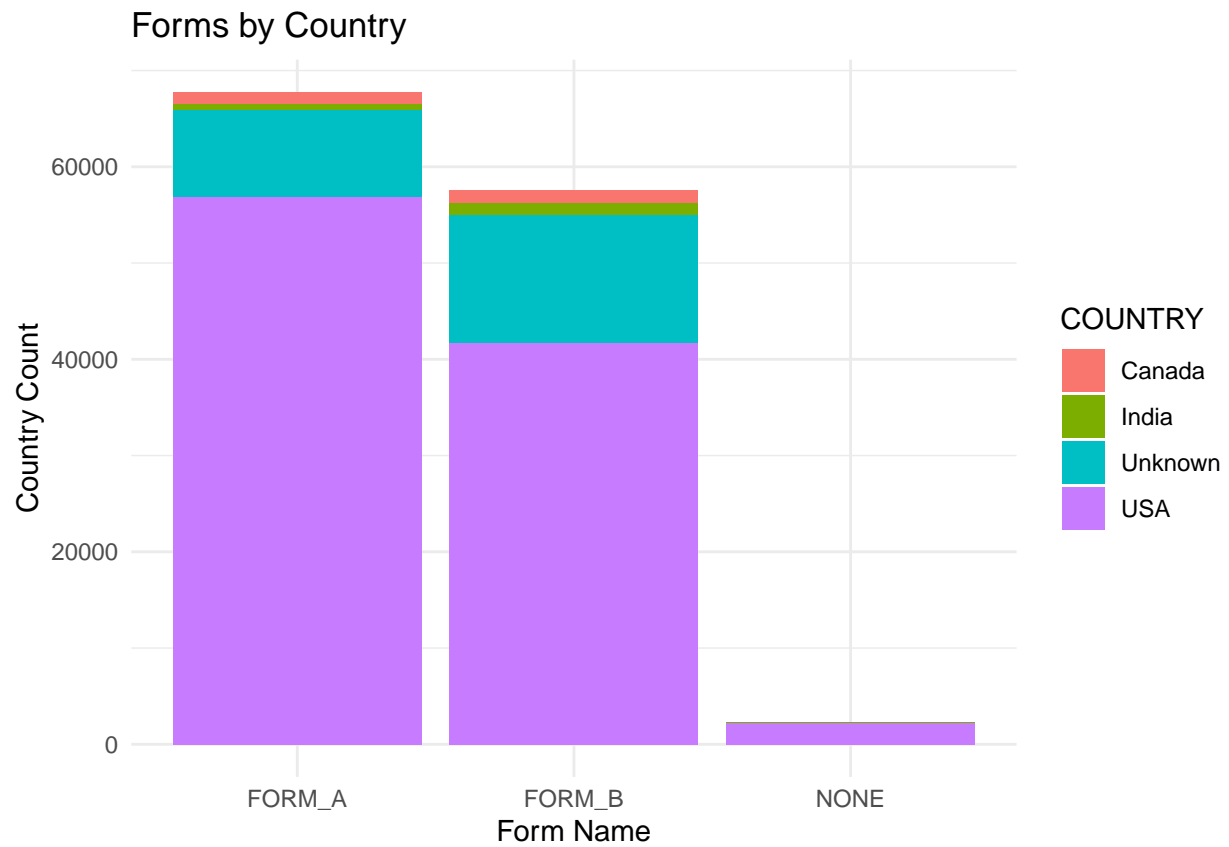
```
## tibble [23,900 x 9] (S3: tbl_df/tbl/data.frame)
## $ UNIQUE_VISITORS: num [1:23900] 8 4 4 4 12 2 12 2 2 50 ...
## $ FORM_STARTS    : num [1:23900] 10 4 6 0 2 2 0 2 2 36 ...
## $ FORM_COMPLETES : num [1:23900] 2 0 2 0 6 0 0 0 0 16 ...
## $ CHANNEL        : Factor w/ 6 levels "Auto Sites","Build & Price Tool",...: 1 4 6 4 5 4 5 5 5 4 ...
## $ FORM_NAME       : Factor w/ 3 levels "FORM_A","FORM_B",...: 1 2 2 3 2 1 3 2 2 2 ...
## $ NEW_DATE        : POSIXct[1:23900], format: "2019-04-07" "2019-04-07" ...
## $ BROWSER         : Factor w/ 5 levels "Chrome","Firefox",...: 1 4 5 3 1 1 4 4 1 1 ...
## $ COUNTRY         : Factor w/ 4 levels "Canada","India",...: 4 4 4 1 4 2 3 3 1 4 ...
## $ VISITOR_TYPE    : Factor w/ 2 levels "New","Returning": 1 2 2 1 1 2 2 1 1 1 ...
```

Plot Analysis

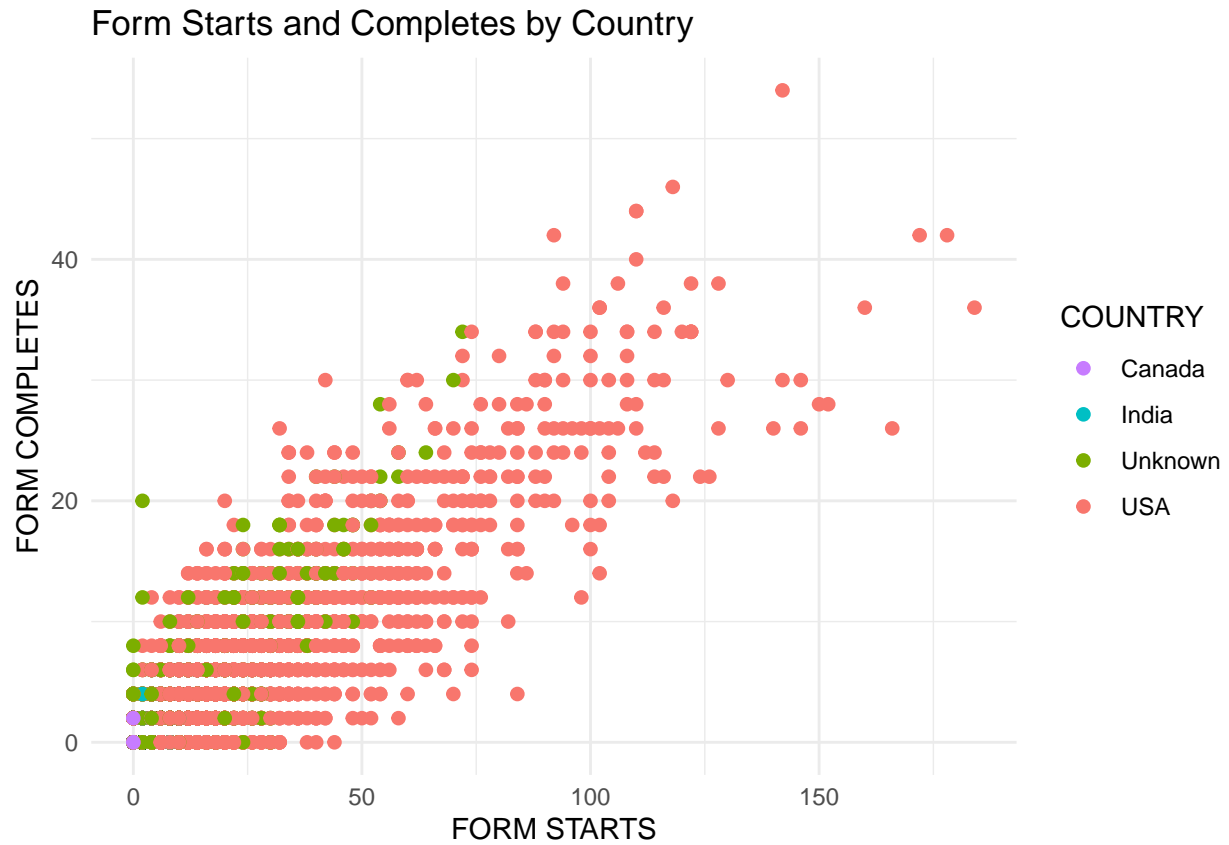
USA has the highest count of unique visitors.



USA makes up the biggest form counts compared to all other countries.



USA makes up most of the data points in form starts and completes

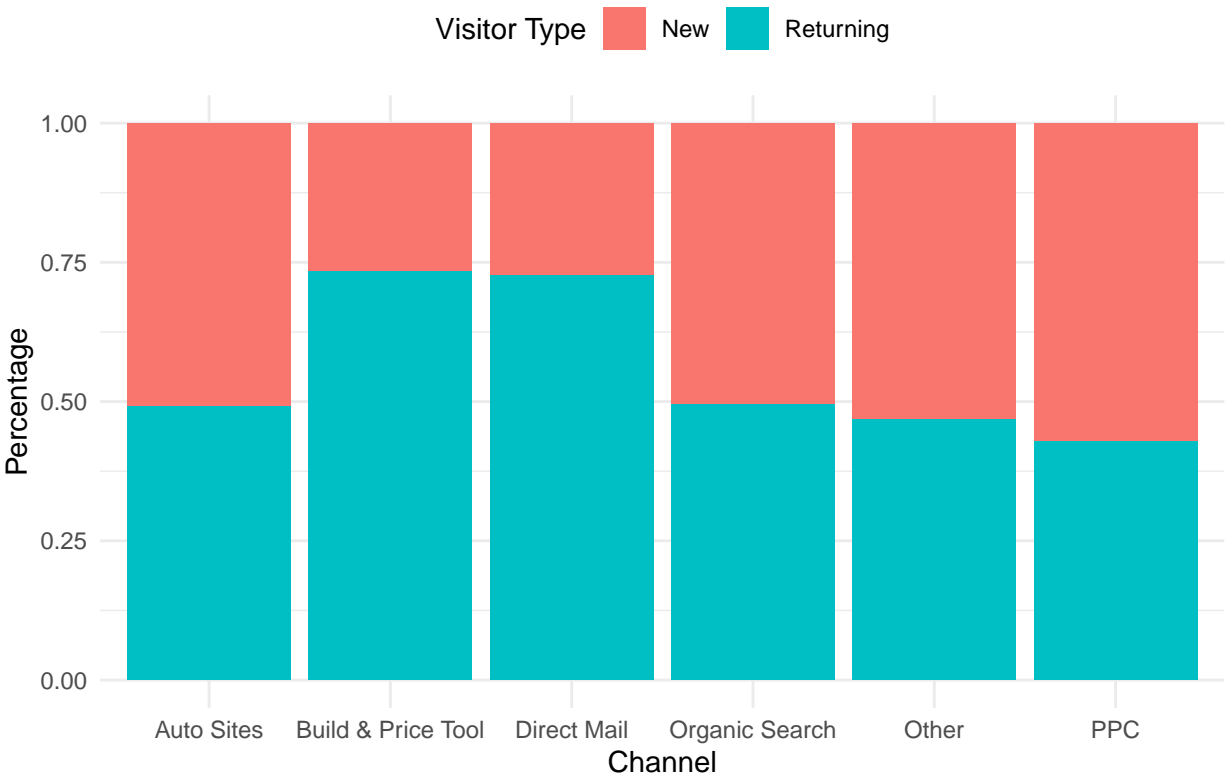


Count of Channel by Visitor Type



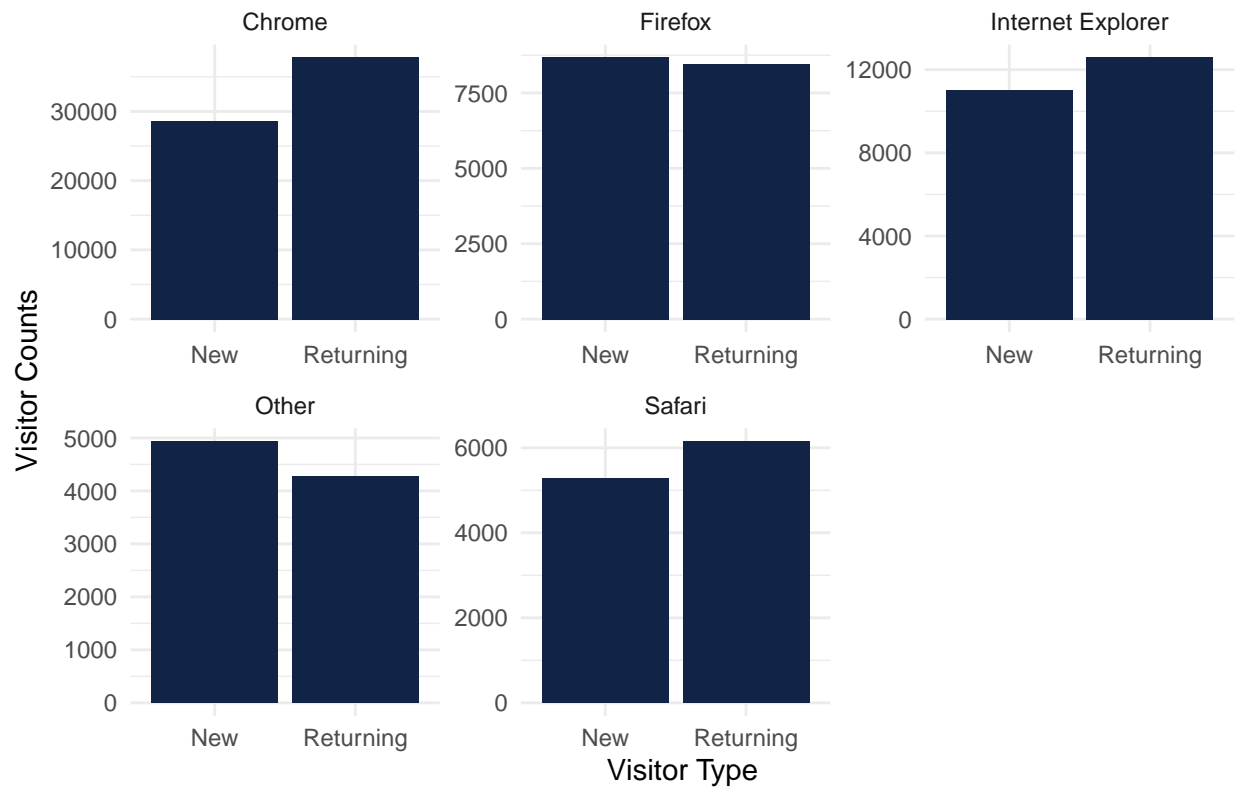
Although overall counts per channel favored Organic Search, a higher percentage of returning visitors 70-74% used Direct Mail and Build & Price Tool.

Visitor Type per Channel Proportions



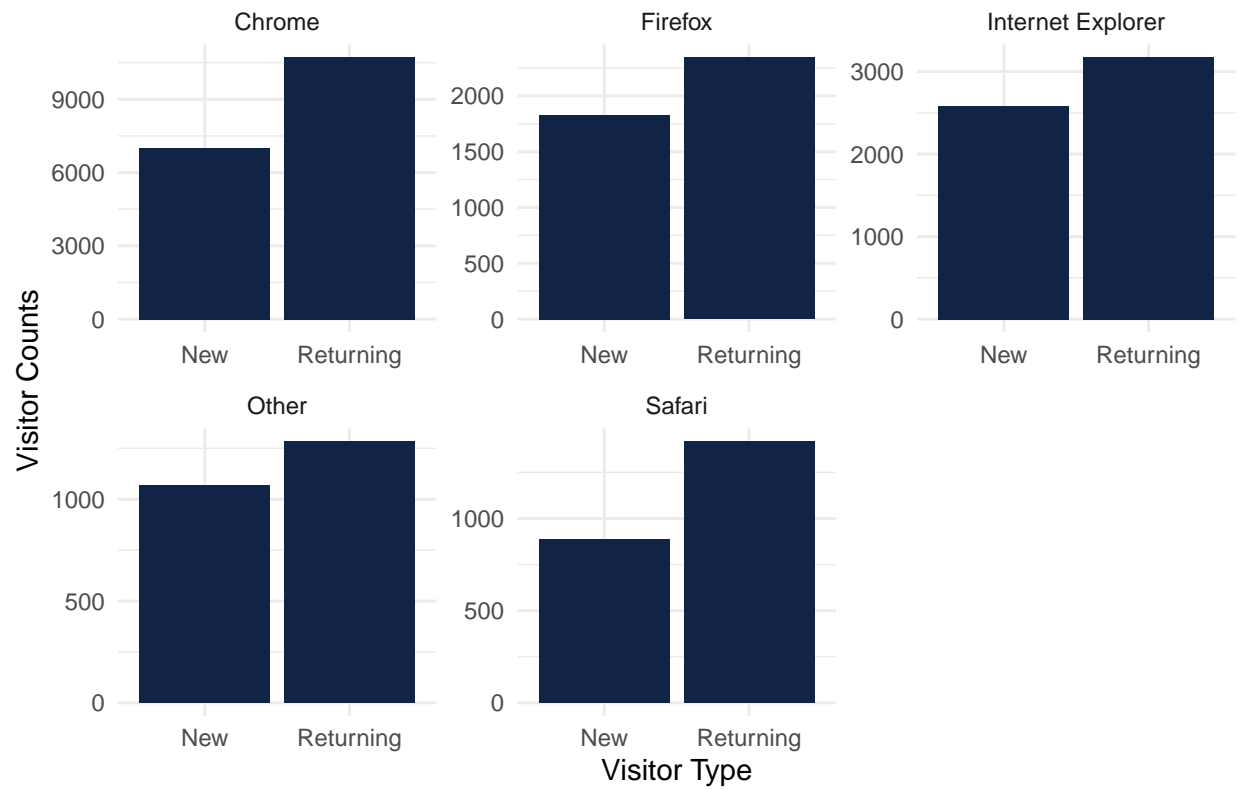
Chrome, Safari, and Internet have more returning customer start forms.

Form Starts by Browser

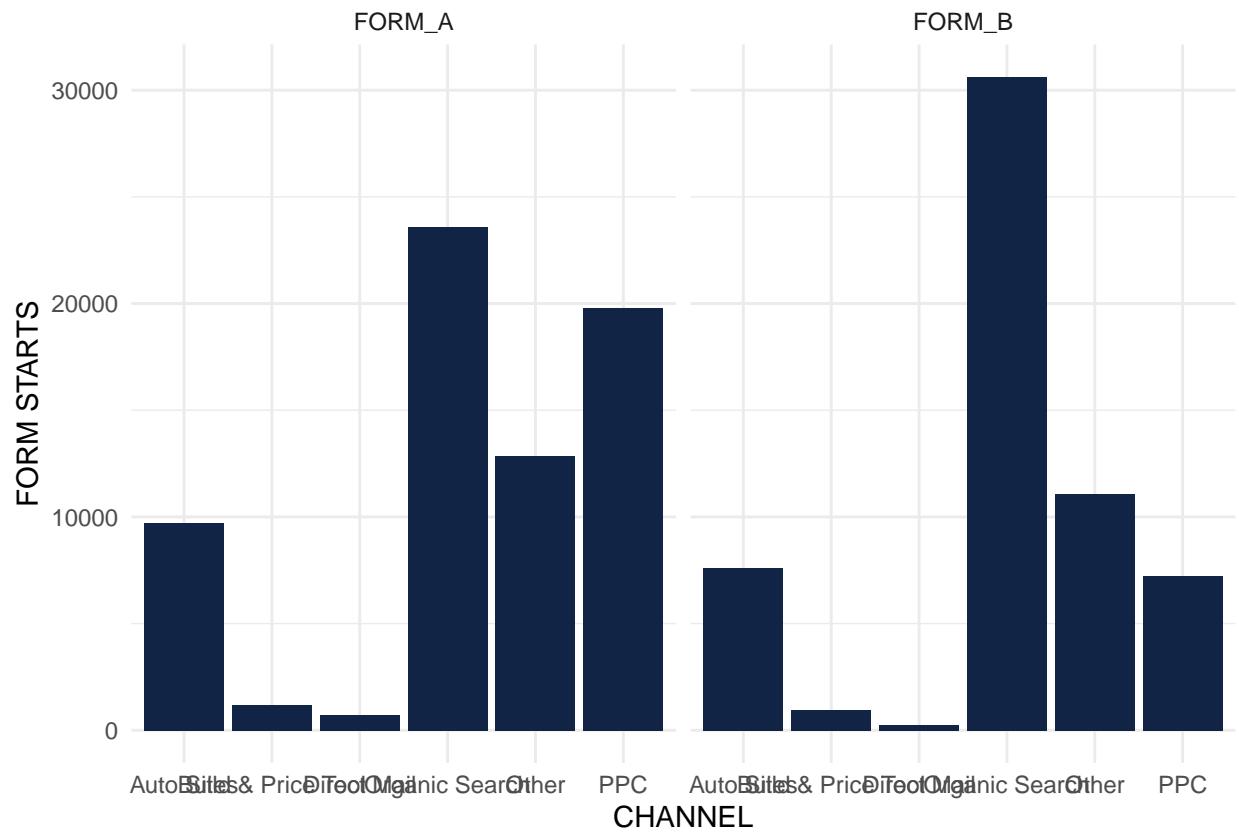


Across all browsers returning customers are more likely to complete the forms.

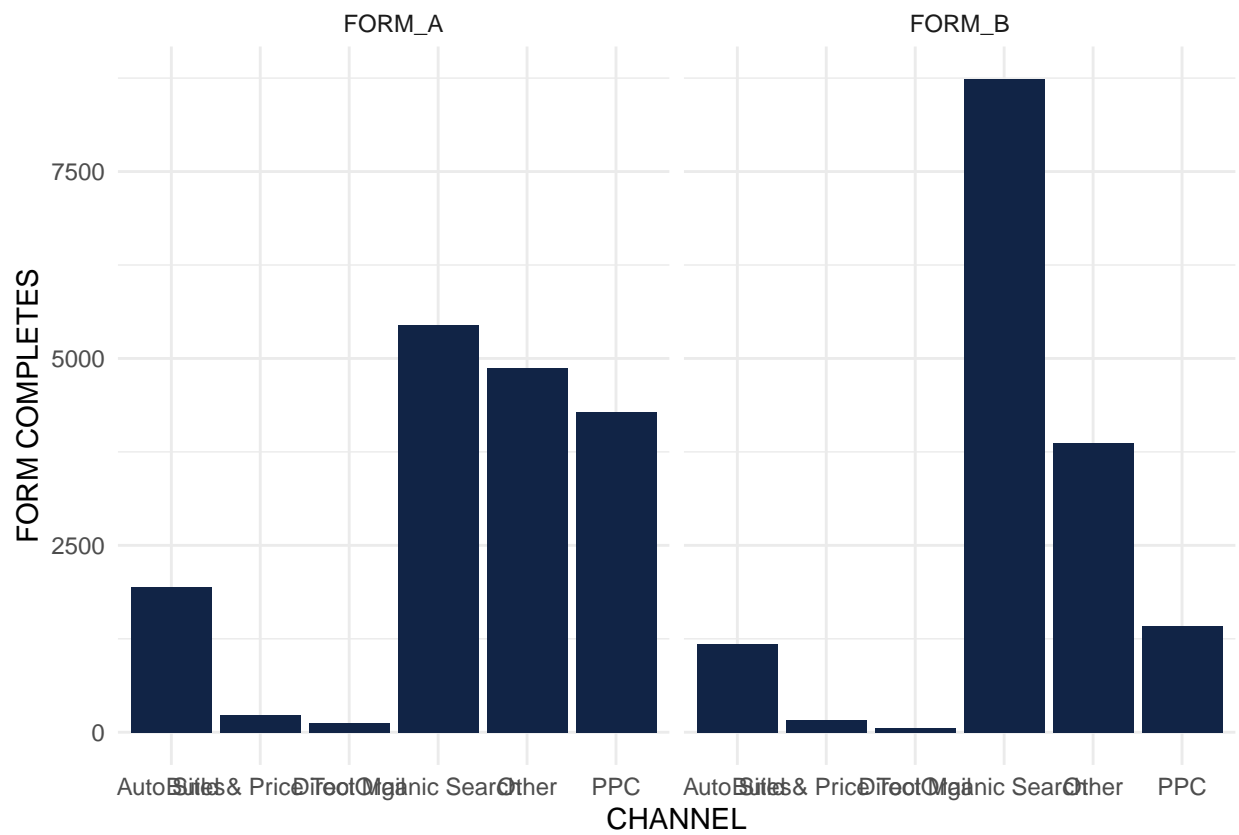
Form completes by Browser



Form B has the most Organic Search starts.



Form B has the most Organic Search completes.



Highest count of visitors appears to be in May with returning visitors dominating across all months.

