

Adidas vs Nike: EDA de productos, precios y reseñas.

Patricio Alvarez Escobar

2023-06-24



1. Introducción.

1.1. Contexto.

Nike y Adidas son dos de las más grandes marcas en la industria de ropa deportiva y calzado deportivo. Ambas marcas representan la excelencia en el desempeño sus productos, en la cual, están en constante innovación e imponiendo las nuevas tendencias en el rubro del vestir deportivo. Ambas marcas han entregado productos icónicos a lo largo de los años, tales como “Jordan´s”, “Nike Air Max”, “Airforce 1”, “Addipure”, “Predator”, y la lista continúa en ambas marcas.

Adidas vs Nike es un debate constante en la industria de la ropa deportiva y el calzado. Este conjunto de datos consiste en la información de los distintos productos en el mercado de estas dos grandes empresas con información significativa. Con este conjunto de datos proporcionado se realizará un Analisis exploratorio de datos (EDA) con el propósito de establecer relaciones competitivas entre ambas marcas.

1.2. Contenido del dataset.

El conjunto de datos consta de 3.268 productos de Nike y Adidas con 12 características de información que incluyen sus calificaciones, descuento, precio de venta, precio de lista, descripción del producto y número de

reseñas. El conjunto de datos se obtuvo del sitio web mundial de datos aportado por datahub y Kaggle. Las ventas mostradas en este dataset pertenecen a 1 mes de ventas.

1.3. Objetivos.

1. Agrupar productos y marcas en función de varios factores.
2. Realizar un análisis exploratorio de los datos y sus distintas variables.
3. Realizar un análisis competitivo de Nike v/s Adidas.

2. Preparación de los datos.

2.1. Cargar librerías.

Antes de iniciar con la preparación del dataset de trabajo, partimos con cargar las librerías a utilizar en el presente proyecto.

```
library(dplyr) # Manipulación de datos.
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
```

```
library(tidyverse) # Manipulación de datos.
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v forcats 1.0.0      v readr 2.1.4
## v ggplot2 3.4.2      v stringr 1.5.0
## v lubridate 1.9.2    v tibble 3.2.1
## v purrr 1.0.1       v tidyr 1.3.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(ggplot2) # Visualización de datos.
library(skimr) # Inspección de la estructura de datos y el EDA.
library(visdat) # Resumen visual de la estructura de los datos.
library(inspectdf) # Inspección de los datos
library(patchwork) # Múltiples plots
```

2.2. Cargar dataset de trabajo.

Se procede a cargar el dataset desde un archivo Excel “.csv” .

```
dataset<- read.csv("Adidas_Nike_dataset.csv")
```

3. Procesamiento de los datos.

Se hace una visualización inicial del dataset por medio de la función `head()` .

```
head(dataset, n = 5)
```

```
##                               Product.Name Product.ID Listing.Price
## 1 Women's adidas Originals NMD_Racer Primeknit Shoes      AH2430      14999
## 2                               Women's adidas Originals Sleek Shoes      G27341      7599
## 3                               Women's adidas Swim Puka Slippers      CM0081       999
## 4 Women's adidas Sport Inspired Questar Ride Shoes      B44832      6999
## 5                               Women's adidas Originals Taekwondo Shoes      D98205      7999
##  Sale.Price Discount                               Brand
## 1         7499        50 Adidas Adidas ORIGINALS
## 2         3799        50          Adidas ORIGINALS
## 3          599        40          Adidas CORE / NEO
## 4         3499        50          Adidas CORE / NEO
## 5         3999        50          Adidas ORIGINALS
##
## 1 Channeling the streamlined look of an '80s racer, these shoes are updated with modern features. The
## 2
## 3
## 4
## 5                               Inspired by modern tech runners, these women's shoes step
## 6                               This design is inspired by vintage Taekwondo styles originally worn to perfect hi
## 7
## 8 Rating Reviews                               Last.Visited
## 9 1      4.8      41 2020-04-13T15:06:14
## 10 2      3.3      24 2020-04-13T15:06:15
## 11 3      2.6      37 2020-04-13T15:06:15
## 12 4      4.1      35 2020-04-13T15:06:15
## 13 5      3.5      72 2020-04-13T15:06:15
```

3.1. Selección de las variables de trabajo.

Hay variables como “Description”, “Product.ID” y “Last.Visited” que no serán parte de este análisis, y que ensucian la visualización de los datos, por lo cual, procedemos a no considerarlas y seleccionamos las otras variables como variables de trabajo.

```
dataset <- dataset %>% select(- Description)
dataset <- dataset %>% select(- Product.ID)
dataset <- dataset %>% select(- Last.Visited)
```

Procedemos a viualizar nuevamente los datos:

Table 1: Table dataset Visualization

Product.Name	Listing.Price	Sale.Price	Discount	Brand	Rating	Reviews
Women's adidas Originals NMD_Racer Primeknit Shoes	14999	7499	50	Adidas Originals	4.8	41
Women's adidas Originals Sleek Shoes	7599	3799	50	Adidas Originals	3.3	24
Women's adidas Swim Puka Slippers	999	599	40	Adidas CORE / NEO	2.6	37
Women's adidas Sport Inspired Questar Ride Shoes	6999	3499	50	Adidas CORE / NEO	4.1	35
Women's adidas Originals Taekwondo Shoes	7999	3999	50	Adidas Originals	3.5	72

3.2. Inspección de la estructura de los datos.

Para visualizar los datos usamos la función `glimpse()` del paquete `dplyr`. Esta función es una versión de la función `print()` y `str()`, en la cual, al aplicarse al dataset, muestra la estructura de los datos, mostrando la mayor cantidad de observaciones posibles. Es una función, que proporciona un resumen mucho mas visual de la estructura de los datos.

Realizamos la visualización de la estructura de los datos:

```
glimpse(dataset)
```

```
> glimpse(dataset)
Rows: 3,268
Columns: 7
$ Product.Name <chr> "Women's adidas Originals NMD_Racer Primeknit Shoes", "Women's adidas Originals Sleek Sh...
$ Listing.Price <int> 14999, 7599, 999, 6999, 7999, 4799, 4799, 999, 5599, 6599, 7599, 6599, 5999, 7599, 7999,...
$ Sale.Price <int> 7499, 3799, 599, 3499, 3999, 1920, 2399, 599, 2799, 3959, 4559, 3959, 2999, 4559, 3999, ...
$ Discount <int> 50, 50, 40, 50, 50, 60, 50, 40, 50, 40, 40, 50, 40, 50, 50, 50, 50, 50, 50, ...
$ Brand <chr> "Adidas Adidas ORIGINALS", "Adidas ORIGINALS", "Adidas CORE / NEO", "Adidas CORE / NEO",...
$ Rating <dbl> 4.8, 3.3, 2.6, 4.1, 3.5, 1.0, 4.4, 2.8, 4.5, 4.0, 2.7, 4.2, 2.6, 3.1, 3.2, 3.1, 1.0, 3.1...
$ Reviews <int> 41, 24, 37, 35, 72, 45, 2, 7, 16, 39, 36, 35, 97, 62, 16, 69, 16, 8, 62, 67, 30, 43, 13,...
```

3.3. Ajuste de variables.

La primera visión de la estructura de los datos nos muestra que hay variables, cuyas observaciones, no representan la clase que debiesen ser. Por lo tanto, es necesario ajustar estas variables.

```
# La variable Brand pasa a ser un factor
dataset$Brand <- as.factor(dataset$Brand)

# La variable Rating pasa a ser a numerico
dataset$Rating <- as.numeric(dataset$Rating)

# La variable Reviews pasa a ser a numerico
dataset$Reviews <- as.numeric(dataset$Reviews)
```

Volvemos a visualizar la estructura de los datos:

```
glimpse(dataset)
```

```
> glimpse(dataset)
Rows: 3,268
Columns: 7
$ Product.Name <chr> "Women's adidas Originals NMD_Racer Primeknit Shoes", "Women's adidas Originals Sleaf Sh...
$ Listing.Price <int> 14999, 7599, 999, 6999, 7999, 4799, 4799, 999, 5599, 6599, 7599, 6599, 5999, 7599, 7999, ...
$ Sale.Price <int> 7499, 3799, 599, 3499, 3999, 1920, 2399, 599, 2799, 3959, 4559, 3959, 2999, 4559, 3999, ...
$ Discount <int> 50, 50, 40, 50, 50, 60, 50, 40, 50, 40, 40, 40, 50, 40, 50, 50, 50, 50, 50, 50, ...
$ Brand <fct> Adidas Adidas ORIGINALS, Adidas ORIGINALS, Adidas CORE / NEO, Adidas CORE / NEO, Adidas ...
$ Rating <dbl> 4.8, 3.3, 2.6, 4.1, 3.5, 1.0, 4.4, 2.8, 4.5, 4.0, 2.7, 4.2, 2.6, 3.1, 3.2, 3.1, 1.0, 3.1...
$ Reviews <dbl> 41, 24, 37, 35, 72, 45, 2, 7, 16, 39, 36, 35, 97, 62, 16, 69, 16, 8, 62, 67, 30, 43, 13,...
```

Ahora la variable Brand, que es una variable categórica, se muestra como “factor” y las variables Rating y Reviews se muestran como integer (es una manera de mostrar números enteros).

3.4. Ajuste de codificación de las observaciones en las variables.

En la variable “Brand” se observa que hay muchas submarcas en el dataset. El objetivo del proyecto es hacer un análisis de las marcas Nike y Adidas, por lo tanto, es necesario que las submarcas las agrupemos en sus marcas principales correspondientes.

```
levels(dataset$Brand) # verificamos los nombres (niveles) a reemplazar
```

```
## [1] "Adidas Adidas ORIGINALS" "Adidas CORE / NEO"
## [3] "Adidas ORIGINALS"       "Adidas SPORT PERFORMANCE"
## [5] "Nike"
```

```
# Establecemos los nuevos nombres
```

```
mynewnames <- c("Adidas Adidas ORIGINALS" = "Adidas",
               "Adidas CORE / NEO" = "Adidas",
               "Adidas ORIGINALS" = "Adidas",
               "Adidas SPORT PERFORMANCE" = "Adidas",
               "Nike" = "Nike")
```

```
# Reemplazamos
```

```
levels(dataset$Brand) <- mynewnames[levels(dataset$Brand)]
```

3.5. Identificar y tratar duplicados.

Procedemos a identificar y eliminar datos duplicados en el dataset. Primero, partimos identificando si hay datos duplicados en el dataset y su cantidad.

```
any(duplicated(dataset))
```

```
## [1] TRUE
```

```
sum(duplicated(dataset))
```

```
## [1] 117
```

Observamos que existen 117 datos duplicados en la tabla, los cuales procedemos a eliminar de la siguiente manera:

```
dataset <- dataset %>% distinct()
```

4. Análisis de los datos.

4.1. Análisis descriptivo de los datos.

En el análisis descriptivo, como su nombre lo indica, describimos las tendencias claves en los datos existentes y observamos las situaciones que conduzcan a llevar a cabo los análisis. Para realizar el análisis descriptivo del dataset de trabajo utilizaremos la función `skim()` del paquete “skimr”. Esta función nos entrega un resumen completo de los datos y su estructura.

```
skim(dataset)
```

Data Summary	
Name	dataset
Number of rows	3151
Number of columns	7
Column type frequency:	
character	1
factor	1
numeric	5
Group variables	None

Variable type: numeric										
skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100	
1 Listing.Price	0	1	7046.	4652.	0	4599	5999	8999	29999	
2 Sale.Price	0	1	5983.	4174.	449	2999	4599	7799	36500	
3 Discount	0	1	27.9	22.4	0	0	40	50	60	
4 Rating	0	1	3.29	1.37	0	2.6	3.6	4.4	5	
5 Reviews	0	1	41.9	31.3	0	12	40	69	223	

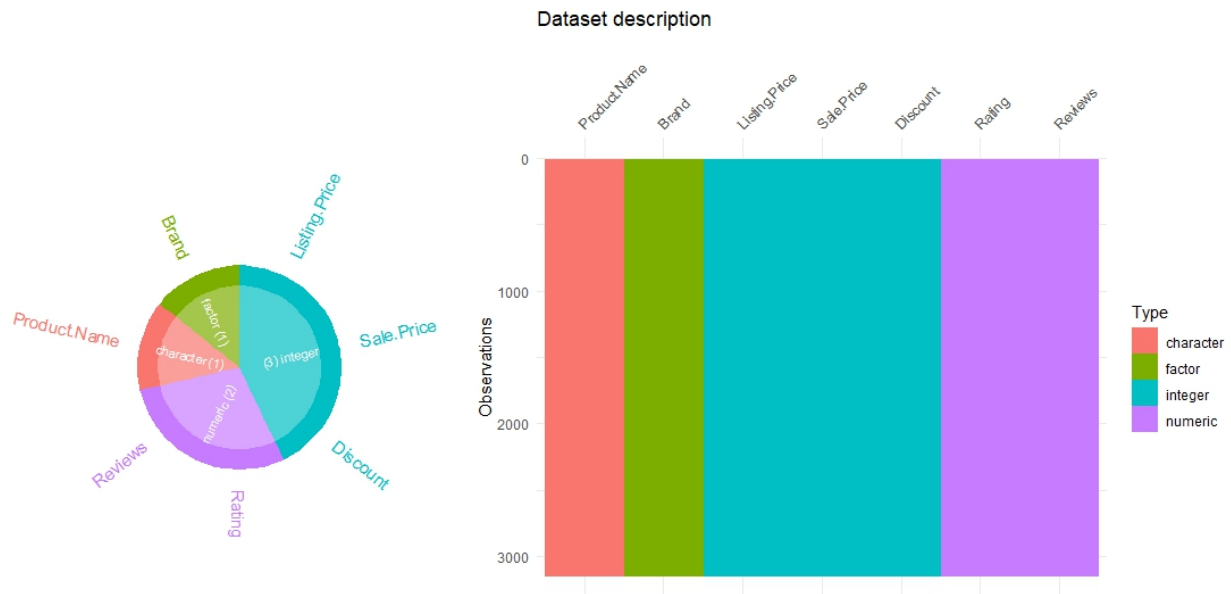
Variable type: character								
skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace	
1 Product.Name	0	1	4	73	0	1531	0	

Los resultados del `skim()` indican que los datos están estructurados por y 3.152 observaciones y 7 variables (columnas). Estas variables se distribuyen en 5 variables numéricas y 2 categóricas. No hay presencia de variables agrupadas.

Al analizar las variables numéricas y su distribución en cuartiles, podemos observar la posible presencia de outliers (valores ubicados en p100).

4.2. Resumen visual de los datos.

El análisis visual de los datos confirma los resultados obtenidos en el análisis descriptivo, en el cual, la mayoría de los datos son numéricos con 7 variables (“precio de lista,” “precio de venta”, “descuentos”, “evaluaciones” y “reseñas”) y existen dos variables categóricas, correspondientes la “marca” y “nombre del producto” (ver figura a continuación).



Por su parte, el análisis de las variables numéricas, confirma lo visto en el análisis del `skim()`, en el cual, se observan datos atípicos (outliers), los cuales pueden afectar el resultado de los análisis (ver figura a continuación).

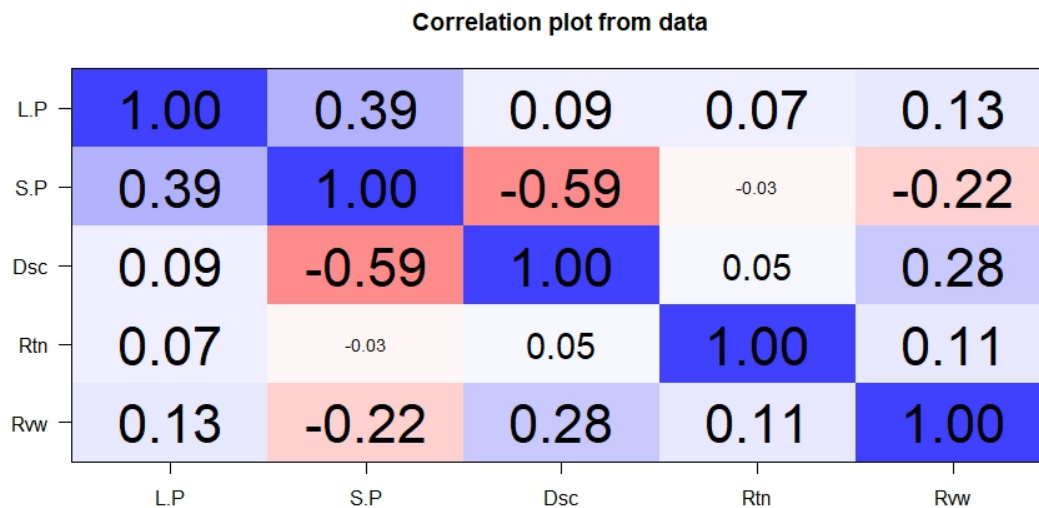
Histograms of numeric columns in df::dataset



4.3. Estableciendo relaciones entre las variables.

Los resultados visuales de correlación muestran que existe relación entre las variables de los datos tales como (ver figura a continuación):

- Existe ligera relación positiva entre el precio de lista con el precio de venta.
- Existe una fuerte relación negativa entre el precio de venta y los descuentos.
- Existe una ligera relación positiva entre los descuentos y las reseñas de los productos.
- Existe una relación negativa entre el precio de venta y las reseñas.



4.4 Análisis de las variables categóricas.

Los resultados muestran que, para estos datos, Adidas presenta casi 5 veces más productos en el mercado que Nike. Los datos indican que la marca Adidas, presenta un total de 2.624 productos diferentes, mientras que Nike solo presenta 527 productos (ver tabla 2).

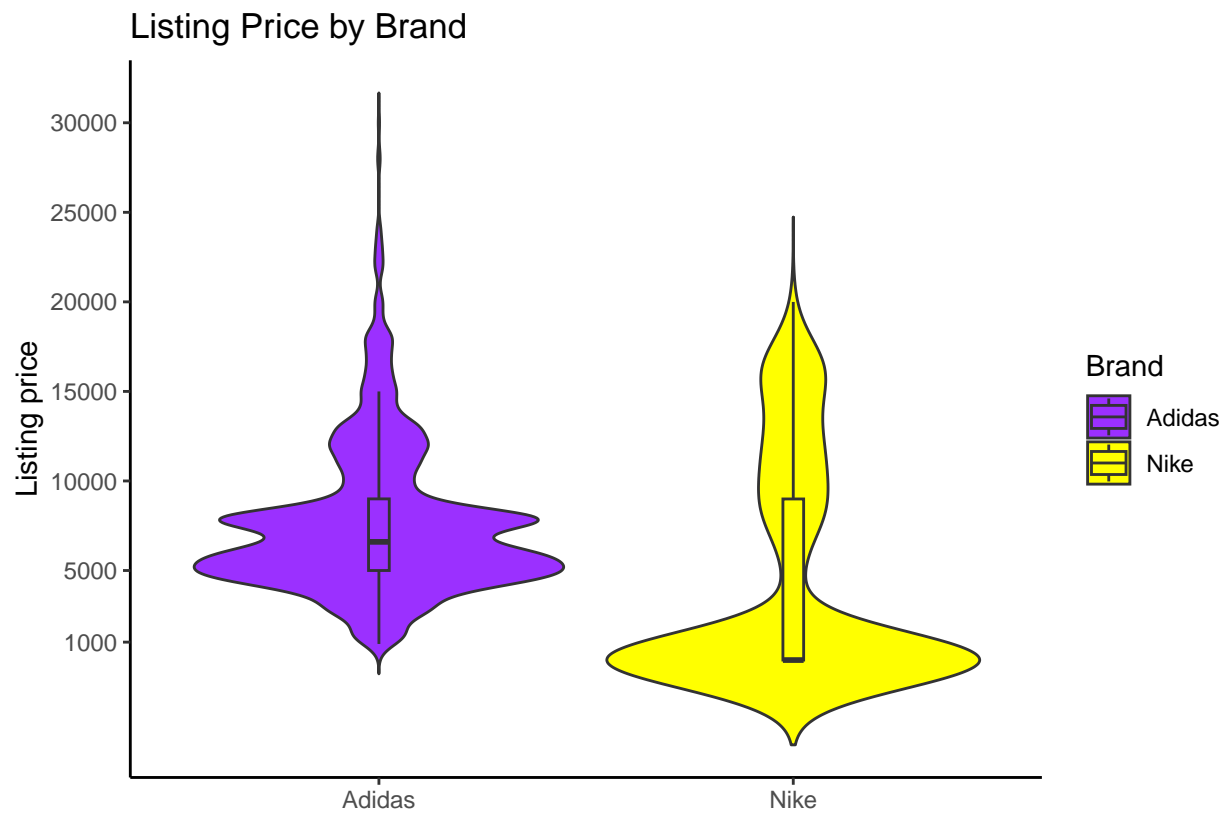
Table 2: Productos disponibles en el mercado

Brand	n
Adidas	2624
Nike	527

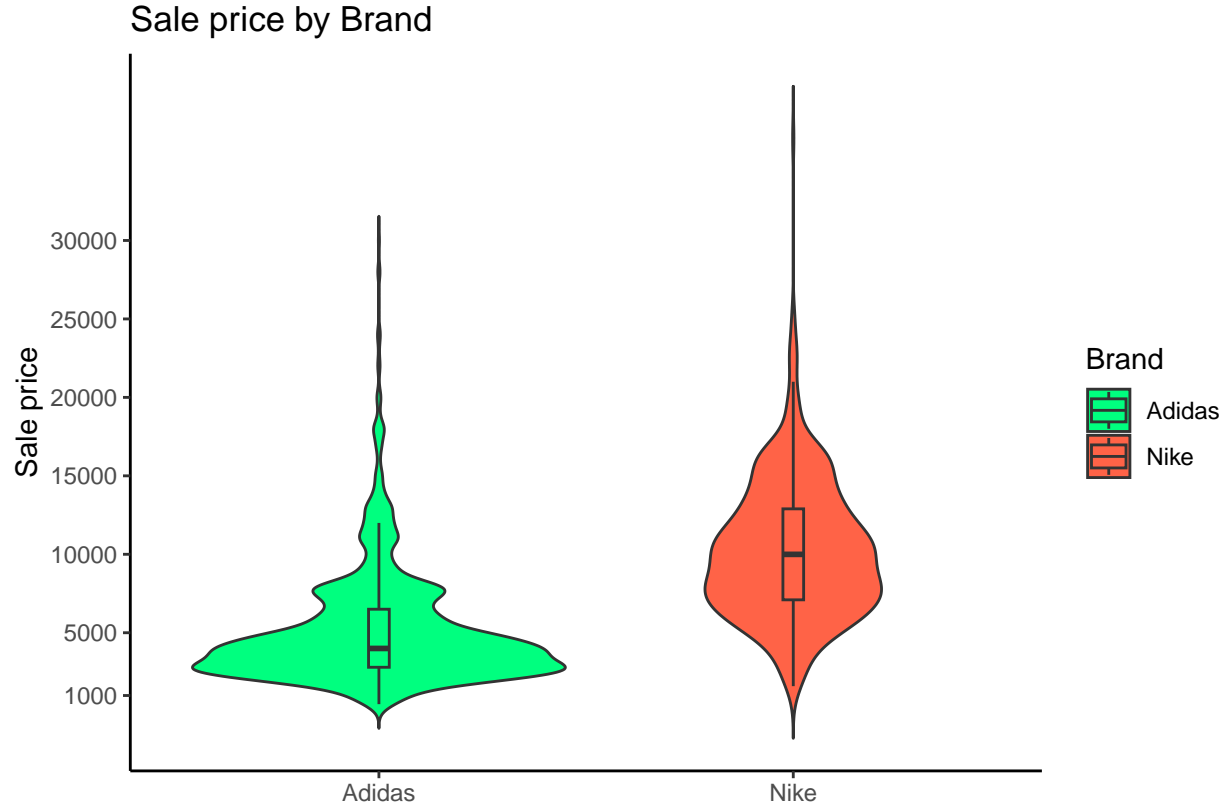
4.5. Analizando las variables numéricas.

4.5.1. Analisis de Precio de lista: Los resultados muestran que, en resumen, Los precios de lista de Adidas son más baratos que los precios de lista de Nike. Los precios de lista de Adidas se distribuyen entre los \$4.000 dólares y los \$9.000 dólares con dos picks, en los cuales, la mayoría de los productos cuestan entre \$5.000 dólares y los \$7.000 dólares. Por su parte, los precios de lista de Nike, que se distribuyen de

una manera más uniforme que los productos adidas, cuestan entre los \$250 dólares y los \$5.000 dólares, con un pick, en el cual, la mayoría de los productos cuestan alrededor de los \$1.000 dólares (ver figura a continuación).



4.5.2. Análisis de Precio de venta: Los resultados del set de datos muestran que, en cuanto al precio de lista, Adidas ofrece sus productos más baratos que Nike. Los precios de venta de Adidas tienen un pick cercano a los \$2.000 dólares, mientras que Nike está más cercano a los \$6.000. Los precios de venta de los productos Nike son mas simétricos comparados con los precios de venta Adidas, que son altamente variables (ver figura a continuación).



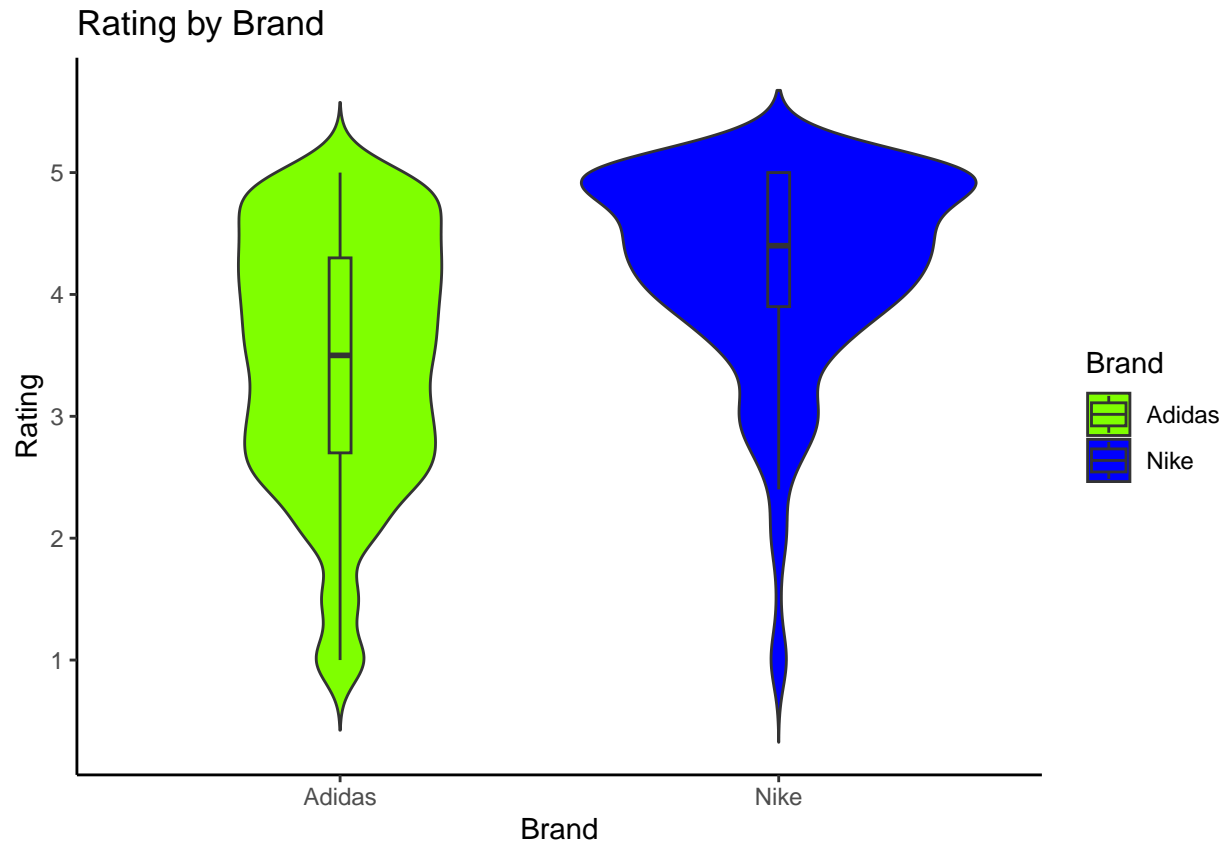
4.5.3 Análisis de las evaluaciones: La tabla 3 muestra un resumen comparativo de Adidas vs Nike en base a los promedios de “Evaluación” de productos, “Precio de lista”, “Precio de venta” y “Reseñas”. Por “Evaluación” se entenderá como la puntuación de satisfacción por parte de los clientes con respecto a un producto. Por su parte, “Reseñas” son el número de comentarios que recibe cada producto por parte de los clientes, ya sean satisfactorios o no satisfactorios.

El análisis de la tabla 3 muestra que, en promedio las evaluaciones de satisfacción de los productos Nike serían mejores que los productos Adidas. Además, comparando el Promedio del número de reseñas vs el promedio de puntuación de las evaluaciones, al parecer Nike tiene mejores evaluaciones en un número menor de reseñas comparado con Adidas.

Table 3: Resumen de promedio de variables Adidas v/s Nike

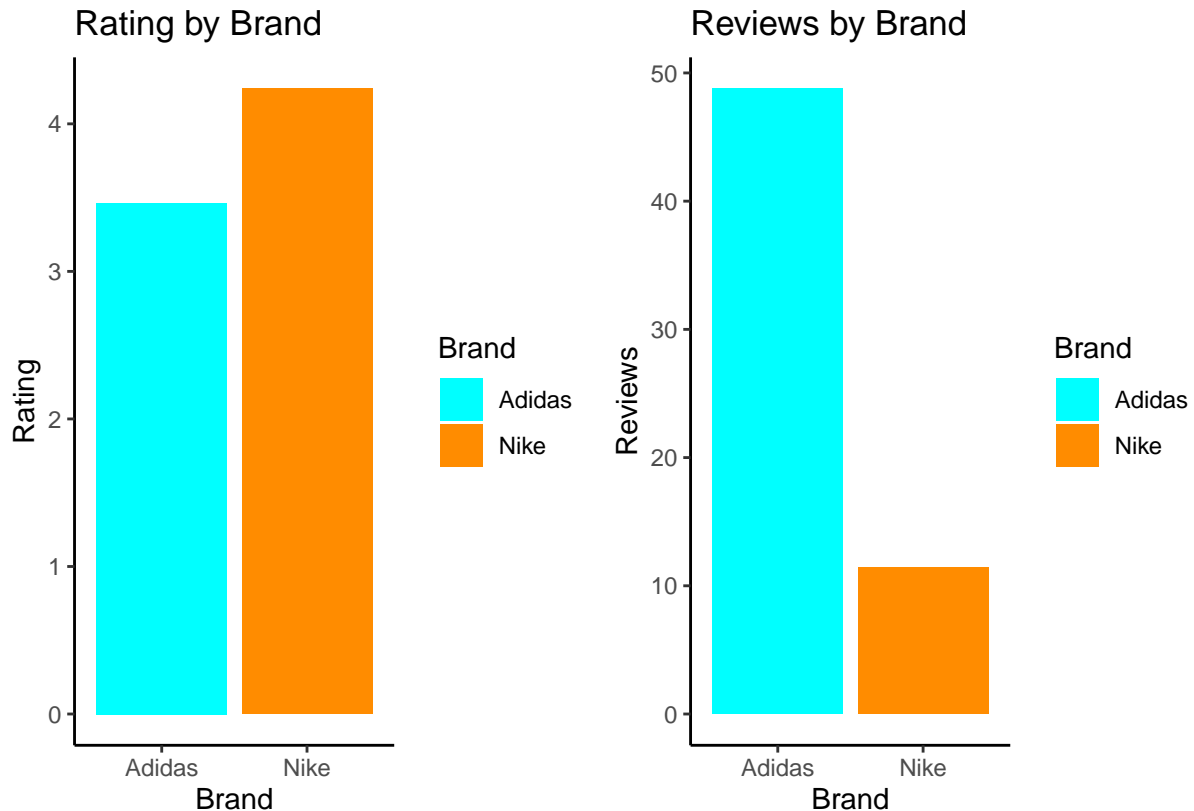
Brand	Mean Rating	Mean Listing.Price	Mean Sale.Price	Mean Reviews
Adidas	3.463622	7607.859	5135.398	48.77695
Nike	4.238827	4659.218	10347.793	11.46369

4.5.4 Análisis de las evaluaciones de producto por marca: El análisis visual de las evaluaciones de los productos confirma lo indicado en la tabla 3. Los resultados indican que la evaluación de productos de Nike son bastante asimétricos y distribuidos a valores altos, esto significa que, para estos datos, Nike tiene mejores evaluaciones que Adidas. Adidas, por su parte, tiene un promedio de evaluaciones de sus productos menor que Nike, pero más uniformes que su contraparte.



4.5.5 Evaluaciones v/s reseñas: El análisis visual de de la contraposición del promedio de reseñas v/s el promedio de puntuación de las evaluaciones por producto de las marcas confirma lo indicado en la tabla 3. Los resultados indican que la evaluación de productos de Nike son, en promedio, más altos que los productos adidas. Por su parte, el promedio de las reseñas por producto son mayores en Adidas que en Nike.

La conclusión de este patrón observado lleva a pensar que, para estos datos, Nike obtiene mejor evaluación de sus productos en una menor cantidad de reseñas por producto que Adidas. Por lo tanto, los resultados hacen pensar que, en promedio, los productos Nike son mucho mejor evaluados que Adidas.



5. Conclusiones.

Las marcas deportivas más importantes del mundo son Adidas y Nike, ambas con una personalidad de marca muy definida y excepcional. Ambas marcas, en su estilo de Marketing, se las han arreglado para liderar el mercado deportivo con sus diversos productos.

En este proyecto se realizó un análisis exploratorio y de visualización de las ventas de estas dos grandes marcas centrándonos en variables competitivas determinadas, tales como: “Productos en el Mercado”, “Precio de Lista”, “Precio de Venta”, “Evaluación de productos” y “Reseñas por producto”. El objetivo de analizar estas variables fue agruparlas para buscar relaciones en base a los distintos factores expuestos y, por medio del análisis exploratorio, buscar las diferencias competitivas entre las marcas Adidas y Nike.

Los resultados de los análisis para éste set de datos indican que:

- Adidas presenta casi 5 veces más productos en el mercado que Nike.
- Los precios de lista de Adidas son más baratos que los precios de lista de Nike.
- En cuanto a los Precios de venta, Adidas ofrece sus productos más baratos que Nike.
- Nike obtiene mejor evaluación de sus productos en una menor cantidad de reseñas por producto que Adidas. Por lo tanto, los resultados hacen pensar que, en promedio, los productos Nike son mucho mejor evaluados que Adidas. haciendo que la experiencia de los productos Nike, para esta base de datos, .
- En resumen, El análisis de las variables competitivas de éste Analisis exploratorio arrojan la conclusión que, para éste set de datos, la experiencia de los productos Nike es mejor que la experiencia de los productos Adidas.