

## Task 6.1 – Sourcing Open Data; Adidas US sales Analysis

### Data Source:

External open-source dataset from Kaggle:

<https://www.kaggle.com/datasets/heemalichaudhari/adidas-sales-dataset>

### Data Collection:

Unfortunately, no details about the collection of this dataset have been provided.

### Contents:

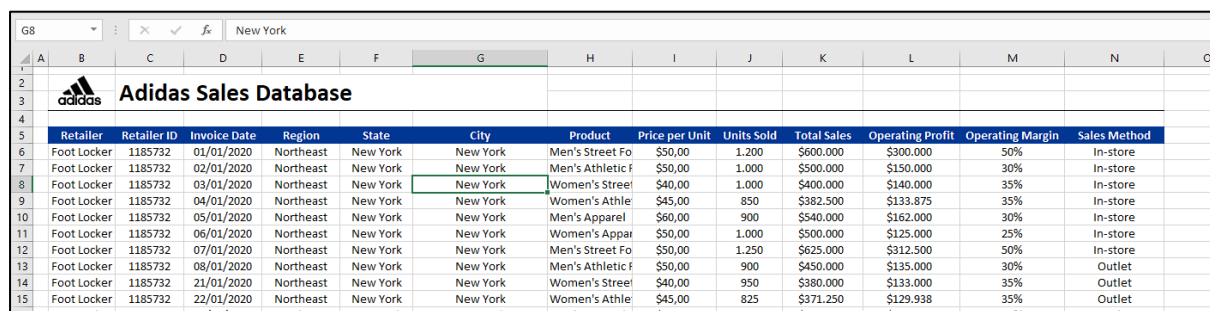
An Adidas sales dataset is a collection of data that includes information on the sales of Adidas products. The dataset includes details such as the number of units sold, the total sales revenue, the location of the sales and the type of product sold.


Data Limitation and Ethics:

### Data Cleaning and Consistency Checks:

Before working with the file, I removed the Adidas logo, the first 4 empty rows and 1 empty column.

Before:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2															
3															
4															
5		Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method	
6		Foot Locker	1185732	01/01/2020	Northeast	New York	New York	Men's Street Fo	\$50.00	1.200	\$600.000	\$300.000	50%	In-store	
7		Foot Locker	1185732	02/01/2020	Northeast	New York	New York	Men's Athletic f	\$50.00	1.000	\$500.000	\$150.000	30%	In-store	
8		Foot Locker	1185732	03/01/2020	Northeast	New York	New York	Women's Street	\$40.00	1.000	\$400.000	\$140.000	35%	In-store	
9		Foot Locker	1185732	04/01/2020	Northeast	New York	New York	Women's Athle	\$45.00	850	\$382.500	\$133.875	35%	In-store	
10		Foot Locker	1185732	05/01/2020	Northeast	New York	New York	Men's Apparel	\$60.00	900	\$540.000	\$162.000	30%	In-store	
11		Foot Locker	1185732	06/01/2020	Northeast	New York	New York	Women's Appar	\$50.00	1.000	\$500.000	\$125.000	25%	In-store	
12		Foot Locker	1185732	07/01/2020	Northeast	New York	New York	Men's Street Fo	\$50.00	1.250	\$625.000	\$312.500	50%	In-store	
13		Foot Locker	1185732	08/01/2020	Northeast	New York	New York	Men's Athletic f	\$50.00	900	\$450.000	\$135.000	30%	Outlet	
14		Foot Locker	1185732	21/01/2020	Northeast	New York	New York	Women's Street	\$40.00	950	\$380.000	\$133.000	35%	Outlet	
15		Foot Locker	1185732	22/01/2020	Northeast	New York	New York	Women's Athle	\$45.00	825	\$371.250	\$129.938	35%	Outlet	
16		Foot Locker	1185732	23/01/2020	Northeast	New York	New York	Men's Apparel	\$60.00	900	\$540.000	\$162.000	30%	Outlet	

Before working with the file, I removed the Adidas logo; the first 4 empty rows and 1 empty column.

After:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method	
2	Foot Locker	1185732	01/01/2020	Northeast	New York	New York	Men's Street Fo	\$50.00	1.200	\$600.000	\$300.000	50%	In-store	
3	Foot Locker	1185732	02/01/2020	Northeast	New York	New York	Men's Athletic f	\$50.00	1.000	\$500.000	\$150.000	30%	In-store	
4	Foot Locker	1185732	03/01/2020	Northeast	New York	New York	Women's Street	\$40.00	1.000	\$400.000	\$140.000	35%	In-store	
5	Foot Locker	1185732	04/01/2020	Northeast	New York	New York	Women's Athle	\$45.00	850	\$382.500	\$133.875	35%	In-store	
6	Foot Locker	1185732	05/01/2020	Northeast	New York	New York	Men's Apparel	\$60.00	900	\$540.000	\$162.000	30%	In-store	
7	Foot Locker	1185732	06/01/2020	Northeast	New York	New York	Women's Appar	\$50.00	1.000	\$500.000	\$125.000	25%	In-store	
8	Foot Locker	1185732	07/01/2020	Northeast	New York	New York	Men's Street Fo	\$50.00	1.250	\$625.000	\$312.500	50%	In-store	
9	Foot Locker	1185732	08/01/2020	Northeast	New York	New York	Men's Athletic f	\$50.00	900	\$450.000	\$135.000	30%	Outlet	
10	Foot Locker	1185732	21/01/2020	Northeast	New York	New York	Women's Street	\$40.00	950	\$380.000	\$133.000	35%	Outlet	
11	Foot Locker	1185732	22/01/2020	Northeast	New York	New York	Women's Athle	\$45.00	825	\$371.250	\$129.938	35%	Outlet	
12	Foot Locker	1185732	23/01/2020	Northeast	New York	New York	Men's Apparel	\$60.00	900	\$540.000	\$162.000	30%	Outlet	
13	Foot Locker	1185732	24/01/2020	Northeast	New York	New York	Women's Appar	\$50.00	1.000	\$500.000	\$125.000	25%	Outlet	

- Checked for duplicates
- Checked for missing values
- Checked for mixed-type columns

### Data Profile:

Dataset contained 9652 rows and 14 columns in total. After cleaning and checking for consistency, the dataset now has 9648 rows and 13 columns.

Column name	Column description	Data Type
Retailer	Name of the retailer where the sale was made	Qualitative, Nominal
Retailer ID	Specific identifying number of the retailer	Quantitative, Discreet
Invoice Date	Date of Sale/Invoice	Qualitative, Ordinal
Region	Region of the sale	Qualitative, Nominal
State	State of the sale	Qualitative, Nominal
City	City of the sale	Qualitative, Nominal
Product	Type of product	Qualitative, Nominal
Price per Unit	Price of the product per unit, in USD	Quantitative, Continuous
Units Sold	Number of units sold	Quantitative, Discreet
Total Sales	Total number of sales, in USD	Quantitative, Continuous
Operating Profit	Total profit for all units sold, in USD	Quantitative, Continuous
Operating Margin	Profit for all units sold, in %	Quantitative, Continuous
Sales Method	The way the products were sold	Qualitative, Nominal

Descriptive analytics of Quantitative data:

1	df_sales.describe()					
	Retailer ID	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin
count	9.648000e+03	9648.000000	9648.000000	9648.000000	9648.000000	9648.000000
mean	1.173850e+06	45.216625	256.930037	93273.437500	34425.244761	0.422991
std	2.636038e+04	14.705397	214.252030	141916.016727	54193.113713	0.097197
min	1.128299e+06	7.000000	0.000000	0.000000	0.000000	0.100000
25%	1.185732e+06	35.000000	106.000000	4254.500000	1921.752500	0.350000
50%	1.185732e+06	45.000000	176.000000	9576.000000	4371.420000	0.410000
75%	1.185732e+06	55.000000	350.000000	150000.000000	52062.500000	0.490000
max	1.197831e+06	110.000000	1275.000000	825000.000000	390000.000000	0.800000

### Data Limitations and Ethics:

One of the limitations of the dataset is the lack of details on collection methods. It takes away the credibility of the dataset. However, for the purpose of this project it is not a critical factor as we are using the results for practice and educational purposes only.

The dataset does not contain any PII of the customers. Only sales of the products are recorded.