King Saud University

Business Administration

College Management Information Systems Department

MIS419

ADIDAS US SALES DATASETS

Section: 54771

Instructor: Dr. Amani Khalaf Samha

Team Members

Student Name	Student ID
Fatimah Almutairy	439201242
Taif Alotaibi	441201843
Aldanah alshubaili	441204296
Reema Alangari	441925239
Deema Almohaya	441202443

Table of content

Introduction

Sales management may better understand where salespeople can improve by using sales analytics to detect, model, comprehend, and predict sales trends and results. The functionality that sales analytic systems offer specifically supports discovery, diagnostic, and predictive exercises that allow the manipulation of parameters, measures, dimensions, or numbers as part of an analytical or planning activity.

A collection of data that contains details on the sales of Adidas products is known as an Adidas sales dataset. This kind of dataset may contain information on the quantity of units sold, the overall revenue from sales, the location of the sales, the kind of product sold, and any other pertinent data.

Adidas sales data can be helpful for several things, including trend analysis, identifying successful items, or marketing initiatives, and formulating sales strategy. Additionally, it can be used to evaluate the efficiency of various marketing or sales channels or to compare Adidas sales to those of rival brands.
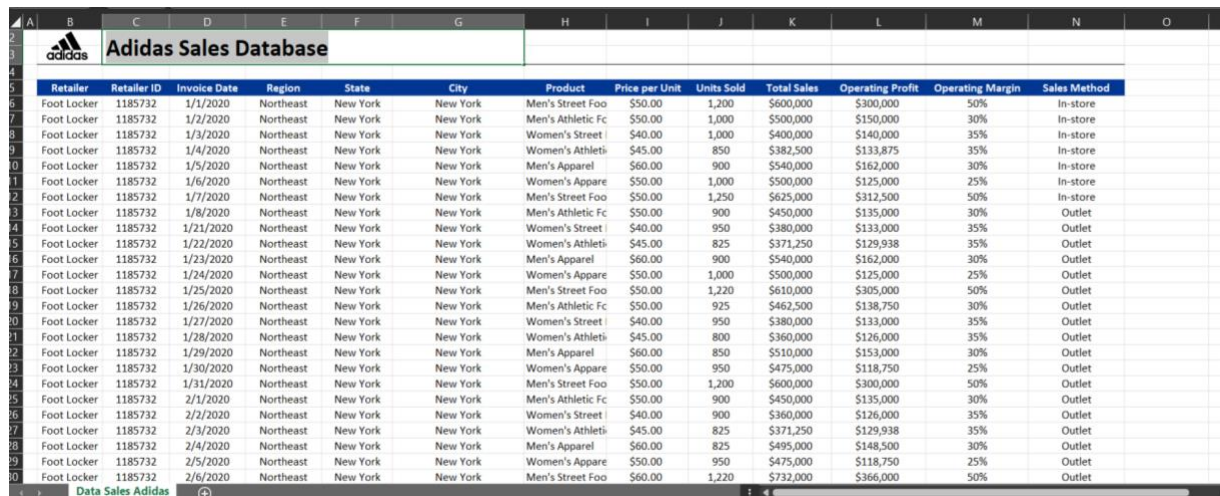
Our dataset " Adidas Sales Database " which is collected from the "Kaggle" website. The dataset contains sales information such as number of products sold, city, price per unit, retailer, invoice date, and sales method...etc.

Microsoft Excel for preprocessing data, Power BI for data visualization, and RapidMiner for pattern discovery are the software that will be used.

The Dataset

The dataset has 9653 rows and 13 columns. This data contains 6 columns of categorical (nominal) type, 6 columns of numerical type, and 1 column of Date-Time type.

The columns are: 'Retailer', 'Retailer ID', 'Invoice Date', 'Region', 'State', City', 'Product', 'Price per Unit', 'Units Sold', 'Total Sales', 'Operating Profit', 'Operating Margin', 'Sales Method'.

Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sales	Operating Profit	Operating Margin	Sales Method
Foot Locker	1185732	1/1/2020	Northeast	New York	New York	Men's Street Foo	$50.00	1,200	$600,000	$300,000	50%	In-store
Foot Locker	1185732	1/2/2020	Northeast	New York	New York	Men's Athletic Fc	$50.00	1,000	$500,000	$150,000	30%	In-store
Foot Locker	1185732	1/3/2020	Northeast	New York	New York	Women's Street	$40.00	1,000	$400,000	$140,000	35%	In-store
Foot Locker	1185732	1/4/2020	Northeast	New York	New York	Women's Athleti	$45.00	850	$382,500	$133,875	35%	In-store
Foot Locker	1185732	1/5/2020	Northeast	New York	New York	Men's Apparel	$60.00	900	$540,000	$162,000	30%	In-store
Foot Locker	1185732	1/6/2020	Northeast	New York	New York	Women's Appare	$50.00	1,000	$500,000	$125,000	25%	In-store
Foot Locker	1185732	1/7/2020	Northeast	New York	New York	Men's Street Foo	$50.00	1,250	$625,000	$312,500	50%	In-store
Foot Locker	1185732	1/8/2020	Northeast	New York	New York	Men's Athletic Fc	$50.00	900	$450,000	$135,000	30%	Outlet
Foot Locker	1185732	1/21/2020	Northeast	New York	New York	Women's Street	$40.00	950	$380,000	$133,000	35%	Outlet
Foot Locker	1185732	1/22/2020	Northeast	New York	New York	Women's Athleti	$45.00	825	$371,250	$129,938	35%	Outlet
Foot Locker	1185732	1/23/2020	Northeast	New York	New York	Men's Apparel	$60.00	900	$540,000	$162,000	30%	Outlet
Foot Locker	1185732	1/24/2020	Northeast	New York	New York	Women's Appare	$50.00	1,000	$500,000	$125,000	25%	Outlet
Foot Locker	1185732	1/25/2020	Northeast	New York	New York	Men's Street Foo	$50.00	1,220	$610,000	$305,000	50%	Outlet
Foot Locker	1185732	1/26/2020	Northeast	New York	New York	Men's Athletic Fc	$50.00	925	$462,500	$138,750	30%	Outlet
Foot Locker	1185732	1/27/2020	Northeast	New York	New York	Women's Street	$40.00	950	$380,000	$133,000	35%	Outlet
Foot Locker	1185732	1/28/2020	Northeast	New York	New York	Women's Athleti	$45.00	800	$360,000	$126,000	35%	Outlet
Foot Locker	1185732	1/29/2020	Northeast	New York	New York	Men's Apparel	$60.00	850	$510,000	$153,000	30%	Outlet
Foot Locker	1185732	1/30/2020	Northeast	New York	New York	Women's Appare	$50.00	950	$475,000	$118,750	25%	Outlet
Foot Locker	1185732	1/31/2020	Northeast	New York	New York	Men's Street Foo	$50.00	1,200	$600,000	$300,000	50%	Outlet
Foot Locker	1185732	2/1/2020	Northeast	New York	New York	Men's Athletic Fc	$50.00	900	$450,000	$135,000	30%	Outlet
Foot Locker	1185732	2/2/2020	Northeast	New York	New York	Women's Street	$40.00	900	$360,000	$126,000	35%	Outlet
Foot Locker	1185732	2/3/2020	Northeast	New York	New York	Women's Athleti	$45.00	825	$371,250	$129,938	35%	Outlet
Foot Locker	1185732	2/4/2020	Northeast	New York	New York	Men's Apparel	$60.00	825	$495,000	$148,500	30%	Outlet
Foot Locker	1185732	2/5/2020	Northeast	New York	New York	Women's Appare	$50.00	950	$475,000	$118,750	25%	Outlet
Foot Locker	1185732	2/6/2020	Northeast	New York	New York	Men's Street Foo	$60.00	1,220	$732,000	$366,000	50%	Outlet

Data Sales Adidas ⊕

To view the entire, excel sheet click here

Data exploration and description

Attribute	Type	Description
Retailer	Nominal	Retailer name
Retailer ID	Numerical	The unique number ID of each retailer
Invoice Date	Date-Time	Dates tamp of each instance
Region	Nominal	Retailer Region
State	Nominal	Retailer State
City	Nominal	Retailer City
Product	Nominal	Types of products
Price per Unit	Numerical	The price by each product
Units Sold	Numerical	The units sold of product by each retailer
Total Sales	Numerical	Total sales of each retailer
Operating Profit	Numerical	Operating profit by retailer
Operating Margin	Numerical	Operating margin by retailer
Sales Method	Nominal	The path that adidas take to reach its end customers in order to sell them the products

Preprocessing techniques

Data preprocessing is the process of transforming raw data into an understandable format and is *the process of gathering, combining, structuring, and organizing data so it can* be used in business intelligence (BI). The quality of the data should be checked before applying machine learning or data mining algorithms. In the preprocessing techniques we used RapidMiner.

1. **discretization using rapid miner**
   the first step after we downloaded our data we have did cleans for everything

2. We add the number of bins we want, and then we choose equal ranges (equal width)

discretization using rapid miner

3. We change the values of the bins will show as (ranges) they are unclear, from TRANSFORM

Transform

Adidas US Sales Datasets (2)

Select columns to transform (hold Shift for selecting a range of columns; Ctrl for (de-)selecting multiple columns; Alt to select all columns of the same type; Ctrl+A for all columns). Make changes and com...

1 column selected

COMMIT TRANSFORMATION   CANCEL   UNDO   SHOW HISTORY

COPY

FILTER

RANGE

SAMPLE

SORT

REPLACE

range3   ✕

very high   ✕

☐ Use regular expressions

✓ APPLY

Retailer	Retailer ID	Invoice Date	Region	State	City	Product	Price per Unit	Units Sold	Total Sa
Category	Number	Date / Time	Category	Category	Category	Category	Number	Category	Number
Sports Direct	1197831	Aug 25, 2021	South	Alabama	Birmingham	Men's Street Foo...	7	low	1568
West Gear	1185732	Sep 12, 2021	West	Utah	Salt Lake City	Women's Street ...	7	low	735
West Gear	1185732	Sep 12, 2021	West	Utah	Salt Lake City	Women's Athleti...	7	low	777
West Gear	1197831	Mar 19, 2021	South	Oklahoma	Oklahoma City	Women's Street ...	7	low	581
Walmart	1197831	Mar 19, 2021	South	Arkansas	Little Rock	Women's Street ...	8	low	896
Sports Direct	1197831	Aug 25, 2021	South	Alabama	Birmingham	Men's Street Foo...	9	low	1827
West Gear	1185732	Sep 12, 2021	West	Utah	Salt Lake City	Women's Street ...	9	low	819
West Gear	1197831	Mar 19, 2021	South	Oklahoma	Oklahoma City	Women's Street ...	9	low	720
Kohl's	1197831	Mar 12, 2021	Midwest	Kansas	Wichita	Women's Street ...	9	low	675
Foot Locker	1185732	Mar 5, 2021	Midwest	Iowa	Des Moines	Women's Street ...	9	low	585

9,648 rows - 13 columns (7 nominal, 5 numerical, 1 date)

values have been successfully replaced.

Data preparation

- ## product

We observed that this attribute contains six values (men's athletics footwear, men's street footwear, men's apparel, women's athletics footwear, women's street footwear and women's apparel). This attribute does not have any missing values, and after checking, we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.
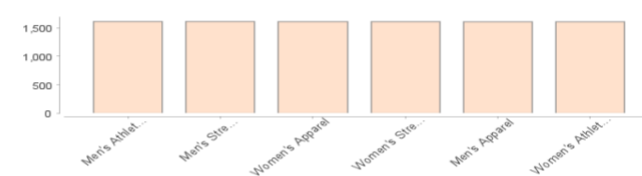
Product

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.06%
Stability: 16.69%
Valid: 83.25%

Top Values

6 Distinct Values:

Value	Count	Percentage
Men's Athletic Footwear	1,610	16.69%
Men's Street Footwear	1,610	16.69%
Women's Street Footwear	1,608	16.67%
Women's Apparel	1,608	16.67%
Men's Apparel	1,606	16.65%

- ## price per unit

We decided to present the price per unit attribute in order manner(ordinal), in order to make it easier and simpler for us. After checking, we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.
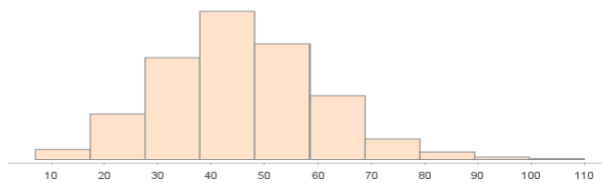
Price per Unit

Summary

Distribution

Number

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.97%
Stability: 5.77%
Valid: 93.25%

Statistics

Name	Value
Minimum	7
Maximum	110.000
Average	45.217
Standard Deviation	14.705

- **sale method**

This attribute consists of values of three categories (online, outlet and in-store). After checking, we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.

Sales Method

Summary

Top Values

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 50.67%
Valid: 49.30%

3 Distinct Values:

Value	Count	Percentage
Online	4,889	50.67%
Outlet	3,019	31.29%
In-store	1,740	18.03%

- **unit sold.**

The attribute consists of a range from (0 to 1275). We decide to convert it from numeric type sections to 3 categorical (discretization), making it easier and simpler for us.
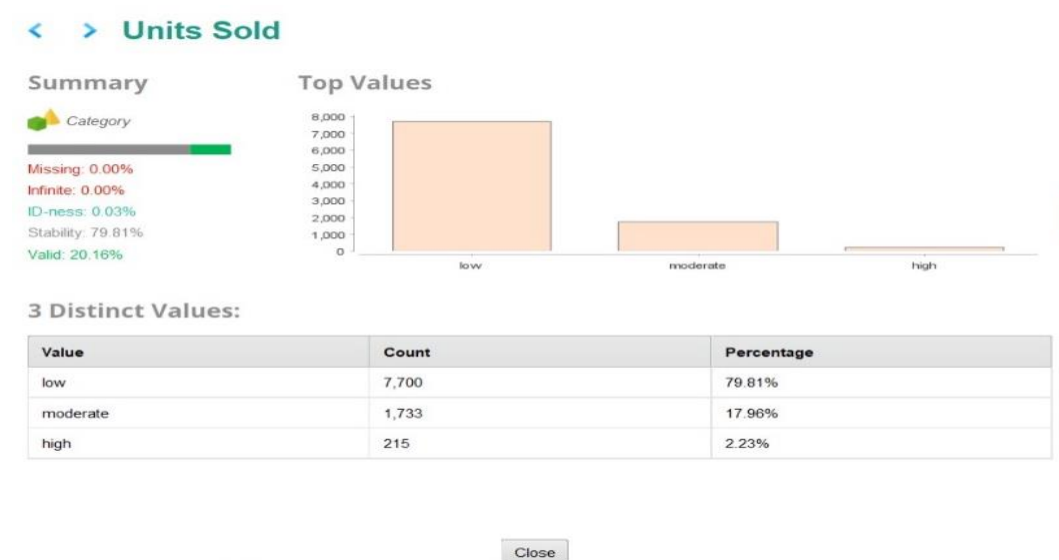
-After the discretization, the categories will be:
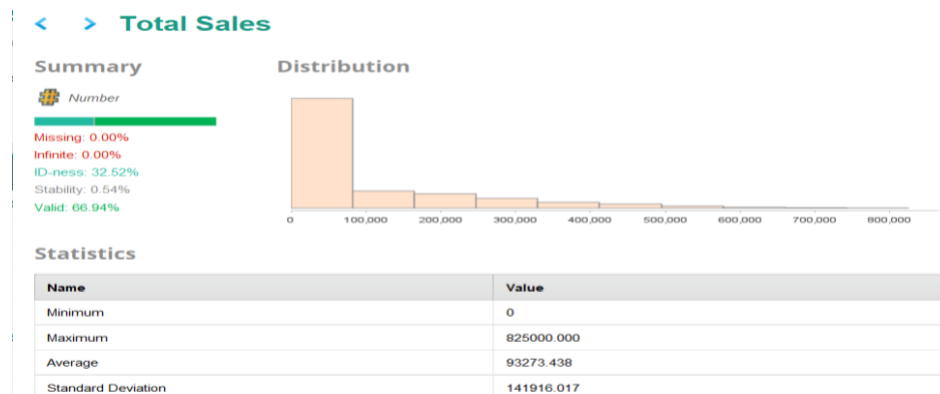
(low): 70.81%

(moderate): 17.96%

(high): 2.23%

We discretized based on equal depth.

Units Sold

Summary — Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.03%
Stability: 79.81%
Valid: 20.16%

3 Distinct Values:

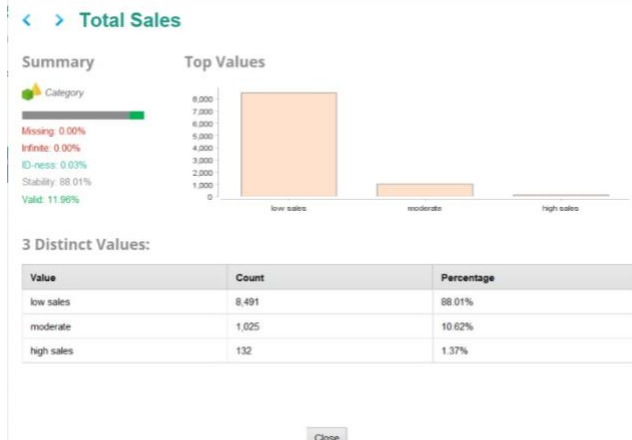| Value	Count	Percentage
low	7,700	79.81%
moderate	1,733	17.96%
high	215	2.23%

- **total sales**

the values are the amounts of total sales. The values are from 0 to 825,000, we decided to convert it from numerical to categorical, so we discretize the Price per unit attribute into 3 equal ranges: low sales, moderate, and high sales.


Total Sales

Summary — Number

Missing: 0.00%
Infinite: 0.00%
ID-ness: 32.52%
Stability: 0.54%
Valid: 66.94%

Statistics

Name	Value
Minimum	0
Maximum	825000.000
Average	93273.438
Standard Deviation	141916.017

here after we change the data into an equal depth

operating profit

the values are the amounts of operating profit. The values are from 0 to 390,000. After checking, we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.
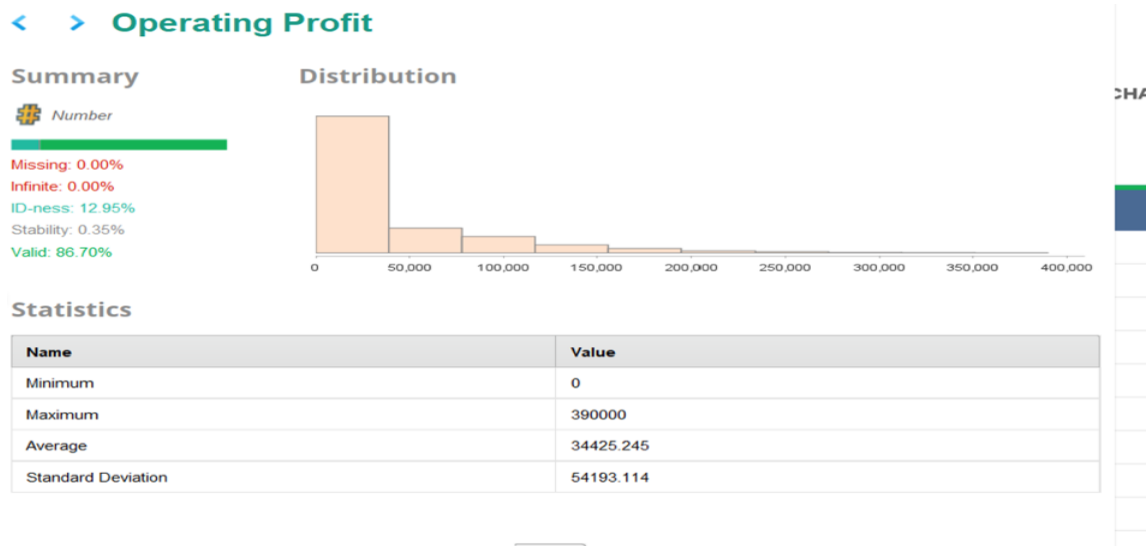

- ### operating margin

the values are the operating margin. The values are from 0.1 to 0.8. After checking, we were able to confirm that there are no Null or NA or blank values. Therefore we

determine that this attribute does not contain any missing values and does not require preprocessing.
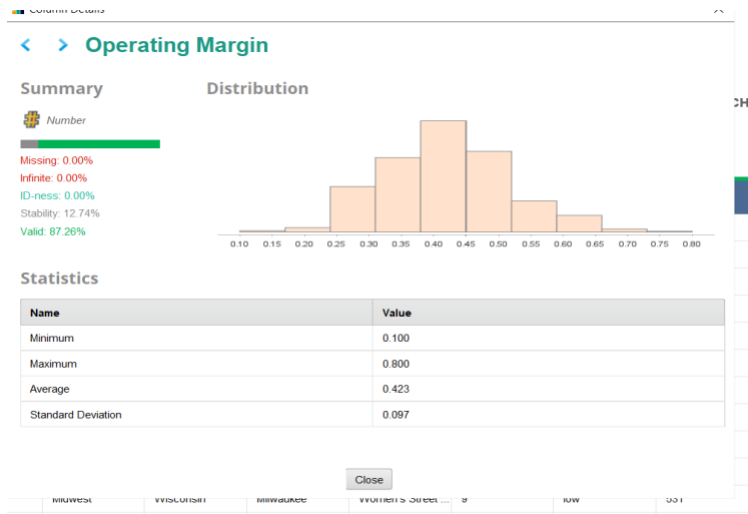
- **invoice date**

the values are dates. The values are from 1 January 2020 to 31 December 2021. After checking, we were able to confirm that there are no Null or NA or blank values. Therefore we determine that this attribute does not contain any missing values and does not require preprocessing.
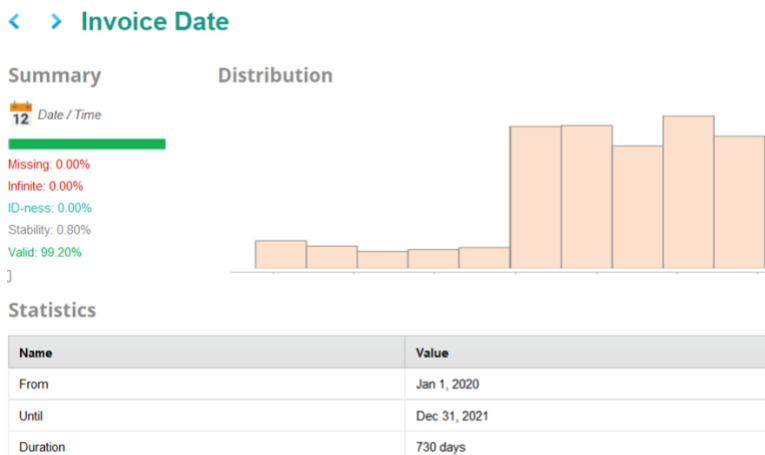

- **Retailer**

    This attribute consists of values of sex categories (Foot locker - west gear - sports direct -kohl's-Amazon-Walmart). After checking through the sort and filter tools,

we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.

‹ › Retailer

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.06%
Stability: 27.33%
Valid: 72.61%
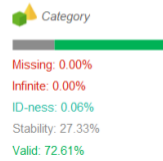
Top Values

6 Distinct Values:

Value	Count	Percentage
Foot Locker	2,637	27.33%
West Gear	2,374	24.61%
Sports Direct	2,032	21.06%
Kohl's	1,030	10.68%
Amazon	949	9.84%

- **State**

  Her in this attribute we have 8 cities (California- Texas- Florida- New York-Alabama - Arizona-Arkansas-Connecticut). After checking through the sort and filter tools, we were able to confirm that there are no Null or NA or blank values. Therefore we determine that this attribute does not contain any missing values and does not require preprocessing.

‹ › State

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.52%
Stability: 4.48%
Valid: 95.00%
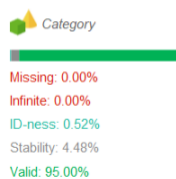
Top Values

50 Distinct Values:

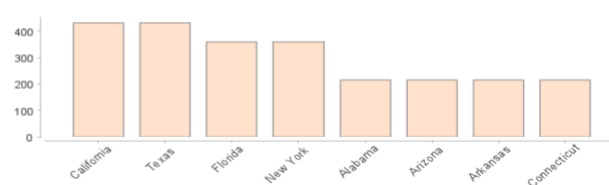| Value	Count	Percentage
California	432	4.48%
Texas	432	4.48%
Florida	360	3.73%
New York	360	3.73%
Utah	216	2.24%

Close

- **City**

Her in this attribute we have 8 cities (California- Texas- Florida- New York-Alabama -Arizona-Arkansas-Connecticut). After checking through the sort and filter tools, we were able to confirm that there are no Null or NA or blank values. Therefore, we determine that this attribute does not contain any missing values and does not require preprocessing.
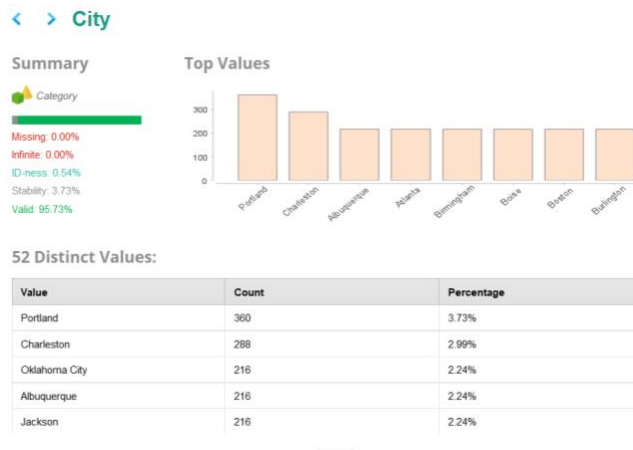

City

Summary

Category

Missing: 0.00%
Infinite: 0.00%
ID-ness: 0.54%
Stability: 3.73%
Valid: 95.73%

52 Distinct Values:

Value	Count	Percentage
Portland	360	3.73%
Charleston	288	2.99%
Oklahoma City	216	2.24%
Albuquerque	216	2.24%
Jackson	216	2.24%

Dataset after Preprocessing:

| Price per U	Total Sales	Operating I	Operating I	Units Sold	Product	Region	Retailer	Sales Meth	State	cluster	id
7.0	low sales	737.0	0.5	low	Men's Street	South	Sports Dire	Outlet	Alabama	cluster_0	8373.0
7.0	low sales	308.7	0.4	low	Women's St	West	West Gear	Outlet	Utah	cluster_0	8610.0
7.0	low sales	357.4	0.5	low	Women's At	West	West Gear	Outlet	Utah	cluster_0	8611.0
7.0	low sales	261.5	0.4	low	Women's St	South	West Gear	Outlet	Oklahoma	cluster_0	9222.0
8.0	low sales	403.2	0.4	low	Women's St	South	Walmart	Outlet	Arkansas	cluster_0	9150.0
9.0	low sales	1077.9	0.6	low	Men's Street	South	Sports Dire	Online	Alabama	cluster_0	4989.0
9.0	low sales	409.5	0.5	low	Women's St	West	West Gear	Online	Utah	cluster_0	5730.0
9.0	low sales	360.0	0.5	low	Women's St	South	West Gear	Online	Oklahoma	cluster_0	6630.0
9.0	low sales	310.5	0.5	low	Women's St	Midwest	Kohl's	Online	Kansas	cluster_0	6702.0
9.0	low sales	234.0	0.4	low	Women's St	Midwest	Foot Locker	Online	Iowa	cluster_0	6918.0
9.0	low sales	233.6	0.4	low	Women's St	Midwest	West Gear	Online	Wisconsin	cluster_0	6990.0
9.0	low sales	158.7	0.4	low	Women's St	Midwest	West Gear	Online	Wisconsin	cluster_0	6996.0
10.0	low sales	31500.0	0.4	moderate	Men's Street	South	Sports Dire	Outlet	Alabama	cluster_0	1101.0
10.0	low sales	12250.0	0.4	low	Women's St	West	West Gear	In-store	Utah	cluster_0	1842.0
10.0	low sales	13000.0	0.4	low	Women's At	West	West Gear	In-store	Utah	cluster_0	1843.0
10.0	low sales	14000.0	0.4	low	Women's St	South	Walmart	Online	Arkansas	cluster_0	2670.0
10.0	low sales	11000.0	0.4	low	Women's St	South	West Gear	Online	Oklahoma	cluster_0	2742.0
10.0	low sales	8750.0	0.4	low	Women's St	Midwest	Kohl's	Online	Kansas	cluster_0	2814.0
10.0	low sales	448.8	0.5	low	Women's At	West	West Gear	Online	Utah	cluster_0	5731.0
10.0	low sales	475.2	0.5	low	Women's St	South	Walmart	Online	Arkansas	cluster_0	6558.0
10.0	low sales	233.2	0.4	low	Women's St	Midwest	Foot Locker	Online	Iowa	cluster_0	6924.0
10.0	low sales	438.7	0.4	low	Women's St	West	West Gear	Outlet	Utah	cluster_0	8580.0

Data Visualization

We've cleaned and formatted the data. Now we want to see the presentation of this data using Power BI to visualize the data and answer the questions we discussed in the introduction.

Total Sales by Retailer
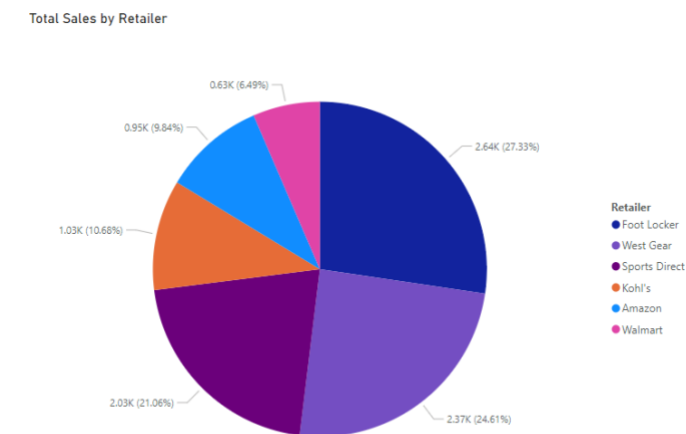
The chart shows top seller is west Gear followed by footlocker; and the lower retailer is Walmart, So Adidas need to work on sales through Walmart.

Product	First Total Sales
Men's Apparel	high sales
Men's Athletic Footwear	high sales
Men's Street Footwear	high sales
Women's Apparel	high sales
Women's Athletic Footwear	low sales
Women's Street Footwear	high sales
Total	**high sales**

After we divided total sales to the three ranges, The chart shows Woman's Athletic Footwear is the lower selling product, so Adidas need to improve or stop selling it.

Count of Units Sold by Sales Method
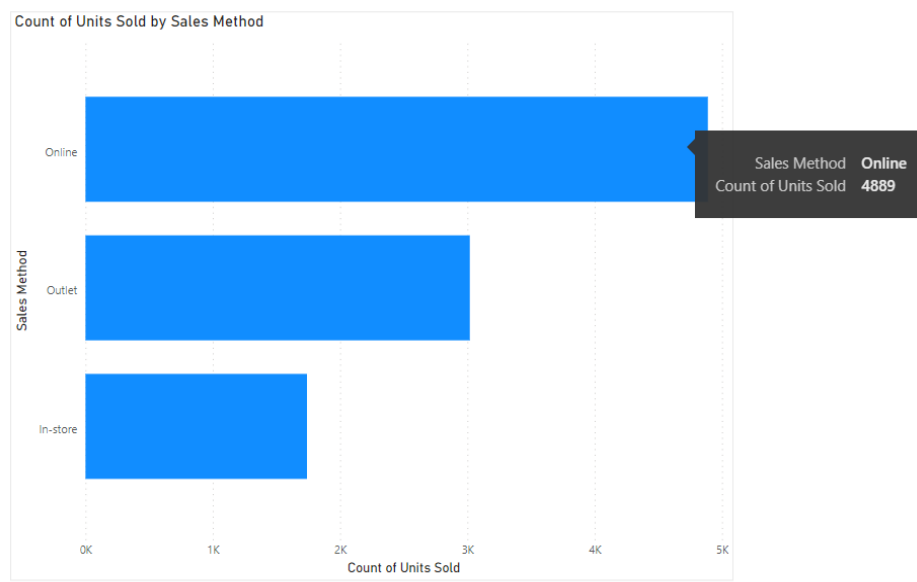
The graph shows Adidas sells more through online and outlet than in-store.

Count of Total Sales by Region
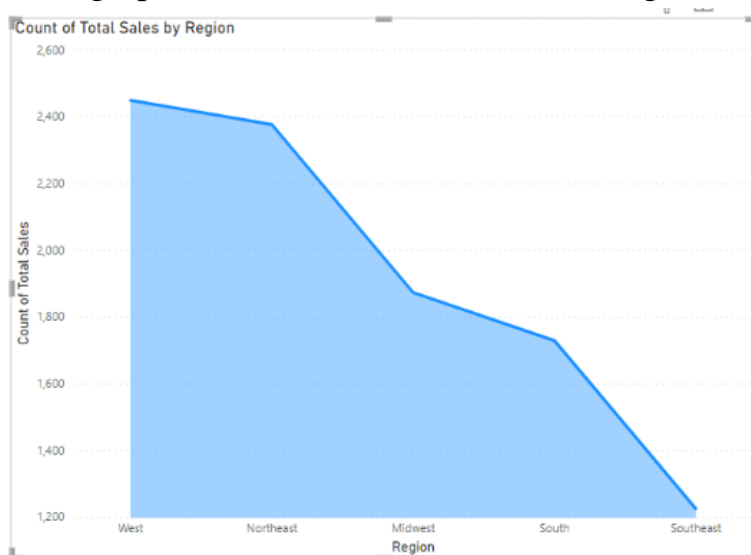
The graph shows the sales in west is the highest than other and contributes highest to the total sales, therefore the Region can make impact on the total sales amount.

Sum of Operating Margin by Sales Method and Product
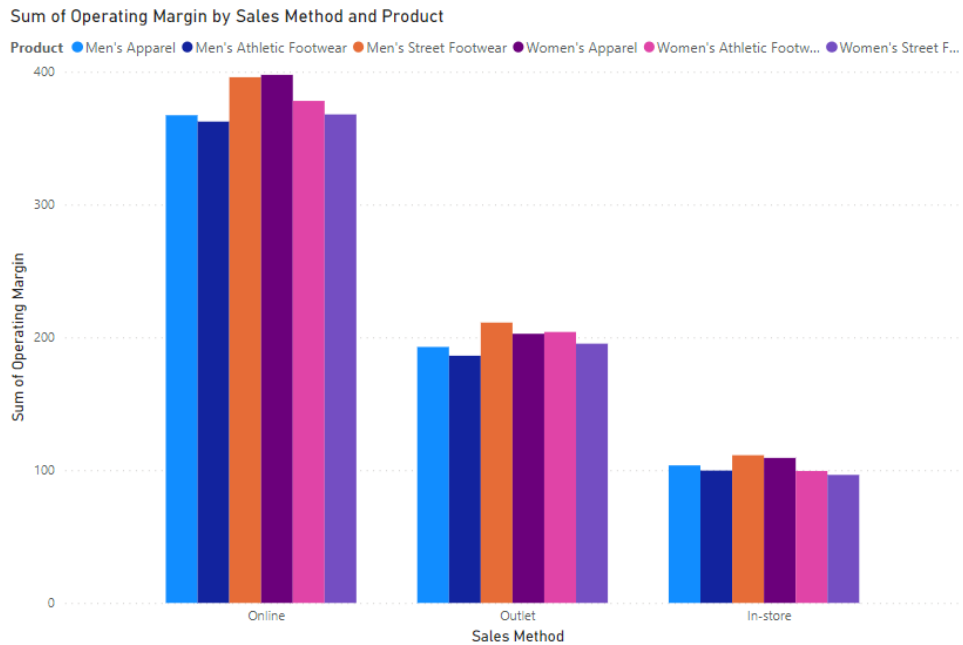
The graph shows the Operating Men's Street footwear earns the highest operating margin and the Online sale has higher operating margin.

Pattern Discovery

We build the model by using the feature (Auto model) in Rapid miner this feature split the dataset by the Cross-validation method.

1-Import the dataset

2-Choose the class label
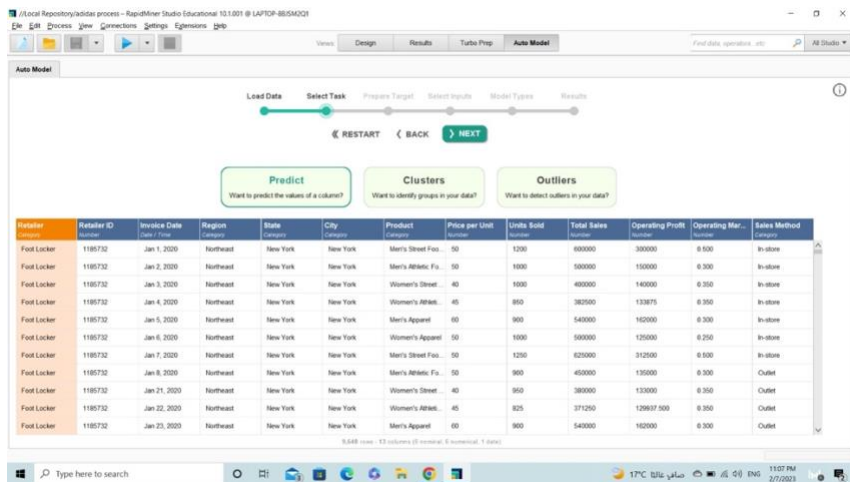
3-Prepare the class label
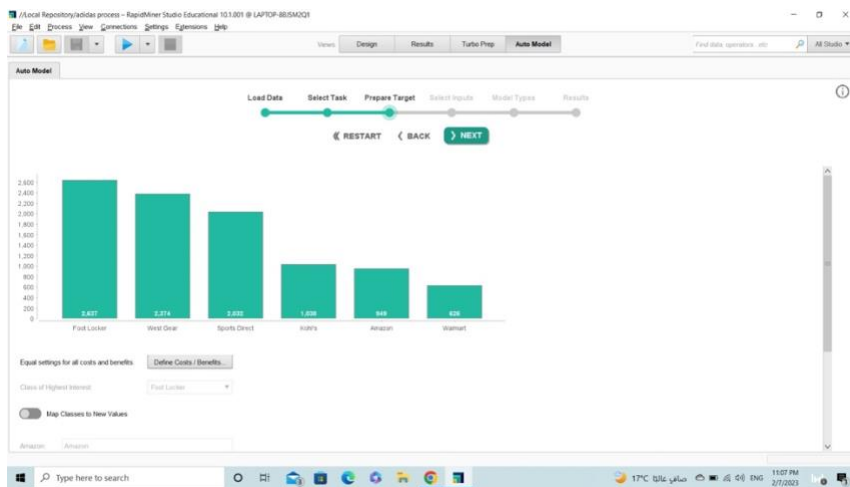
4-Select inputs (attributes)
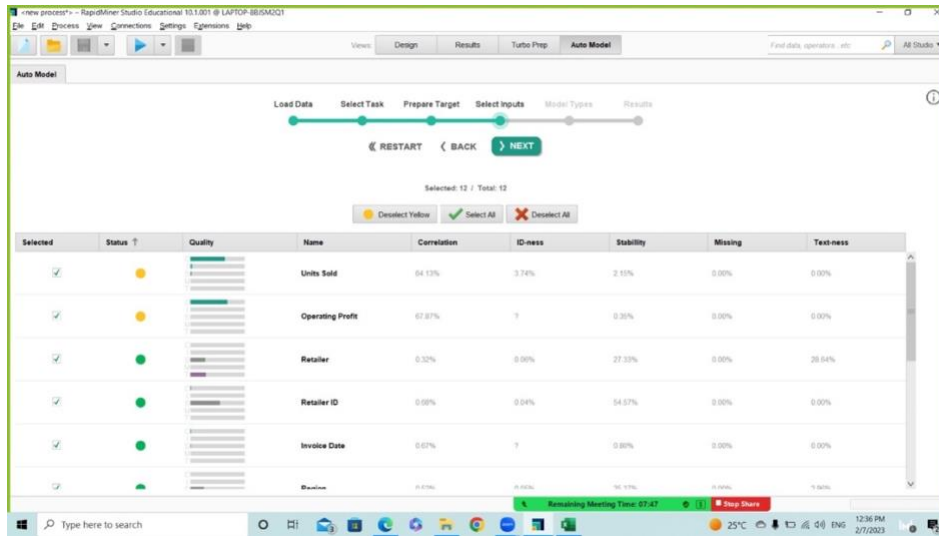
5-Choosing model

6-Implementing models

k-Means - Summary

Number of Clusters: 2

Cluster 0 7,341

Total Sales is on average **73.23%** smaller, **Operating Profit** is on average **72.38%** smaller, **Units Sold** is on average **38.80%** smaller

Cluster 1 2,307

Total Sales is on average **233.02%** larger, **Operating Profit** is on average **230.33%** larger, **Units Sold** is on average **123.47%** larger

k-Means - Heat Map

k-Means - Scatter Plot

Focus on Cluster:
Cluster 0

Jitter Amount:
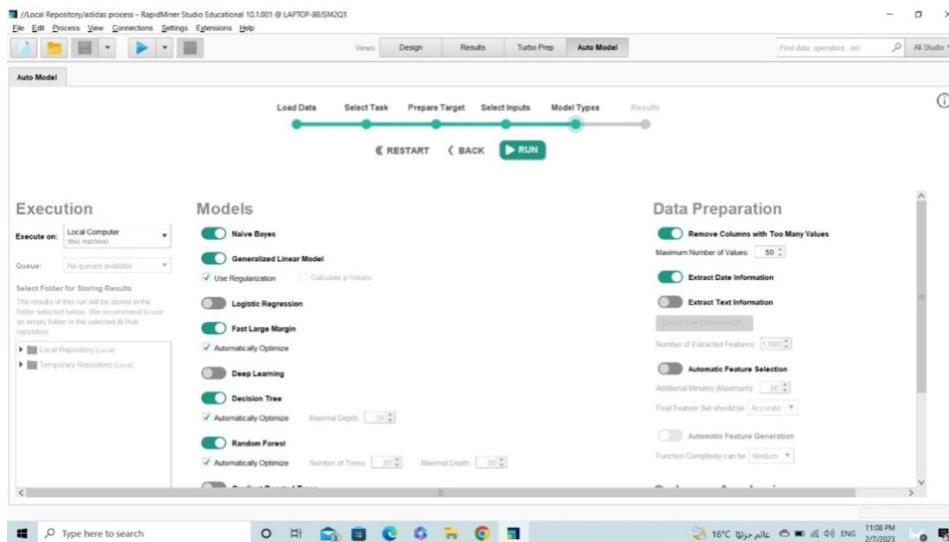
Performance Measurement

Decision Tree - Performance

Performances

Criterion	Value	Standard Deviation
Accuracy	92.0%	± 0.5%
Classification Error	8.0%	± 0.5%

Confusion Matrix

	true Foot Locker	true Walmart	true Sports Direct	true West Gear	true Kohl's	true Amazon	class precision
pred. Foot Locker	687	2	20	22	3	12	92.09%
pred. Walmart	1	122	1	3	0	0	96.06%
pred. Sports Direct	30	47	555	9	7	6	84.86%
pred. West Gear	17	6	0	642	1	5	95.68%
pred. Kohl's	2	0	5	2	283	0	96.92%
pred. Amazon	14	0	2	4	0	248	92.54%
class recall	91.48%	68.93%	95.20%	94.13%	96.26%	91.51%	

Because our confusion matrix became very large, we decide to divide it into three matrixes as shown below.

Load Data   Select Task   Prepare Target   Select Inputs   Model Types   Results

« RESTART   ‹ BACK   ⊃• OPEN PROCESS   ⤓ EXPORT

Results

Simulator
Performance
Predictions
Production Model

▸ Generalized Linear Model

▸ Fast Large Margin

▾ Decision Tree
  Model
  Weights
  Simulator
  Performance
  Optimal Parameters
  Predictions
  Production Model

SAVE RESULTS

Decision Tree - Performance

Performances

Criterion	Value	Standard Deviation
Accuracy	92.0%	± 0.5%
Classification Error	8.0%	± 0.5%

Confusion Matrix

	true Foot Locker	true Walmart	true Sports Direct	true West Gear	true Kohl's	true Amazon	class precision
pred. Foot Locker	687 TP	2 FN	20	22	3	12	92.09%
pred. Walmart	1 FP	122 TN	1	3	0	0	96.06%
pred. Sports Direct	30	47	555	9	7	6	84.86%
pred. West Gear	17	6	0	642	1	5	95.68%
pred. Kohl's	2	0	5	2	283	0	96.92%
pred. Amazon	14	0	2	4	0	248	92.54%
class recall	91.48%	68.93%	95.20%	94.13%	96.26%	91.51%	

As it's shown the percentages of the accuracy, precision and recall are very high.

Conclusion

Finally, the project was an enriching journey rich in knowledge and pleasure that refined our skills and opened wide horizons for us.

This has improved our ability to work with datasets, understand them, analyze them, and extract insightful patterns from process known as data mining.

References and resources:

- **https://www.kaggle.com**

- **Rapid Miner**

- **Microsoft Excel**