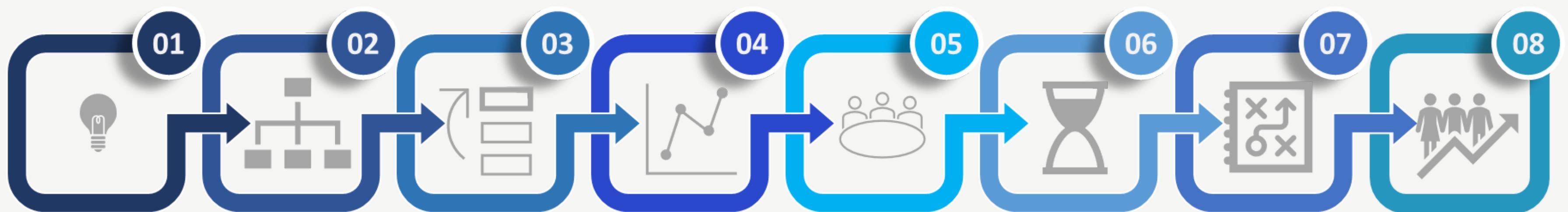


WALMART SALES FORECASTING



STORE SALES FORECASTING



Problem
Definition

Problem
Structure

Prioritize
Levers

Determine
Analysis

Develop
Work plan

Conduct
Analysis

Synthesize
Result

Develop
Recommendation

Problem Definition

Basic Question to Be Resolved

Initial Situation

Success Criteria

Project Scope

Decision Makers

Constraints

Basic Question to Be Resolved

Is Walmart able to improve
customer satisfaction and ensure
Product availability?



Initial Situation

Walmart weekly sales are decreasing
by about 29% after holidays due to
"Out Of Stock" Products



Success Criteria

Predict weekly Sales accurately to optimize inventory management and maximize revenue by 20% in 6 months



Project Scope

United States

Decision Makers

Board of
Directors

Chief Executive
Officer (CEO)

Regional and
Divisional
Managers

Senior
Executives

Store
Managers

Constraints

Market
Saturation

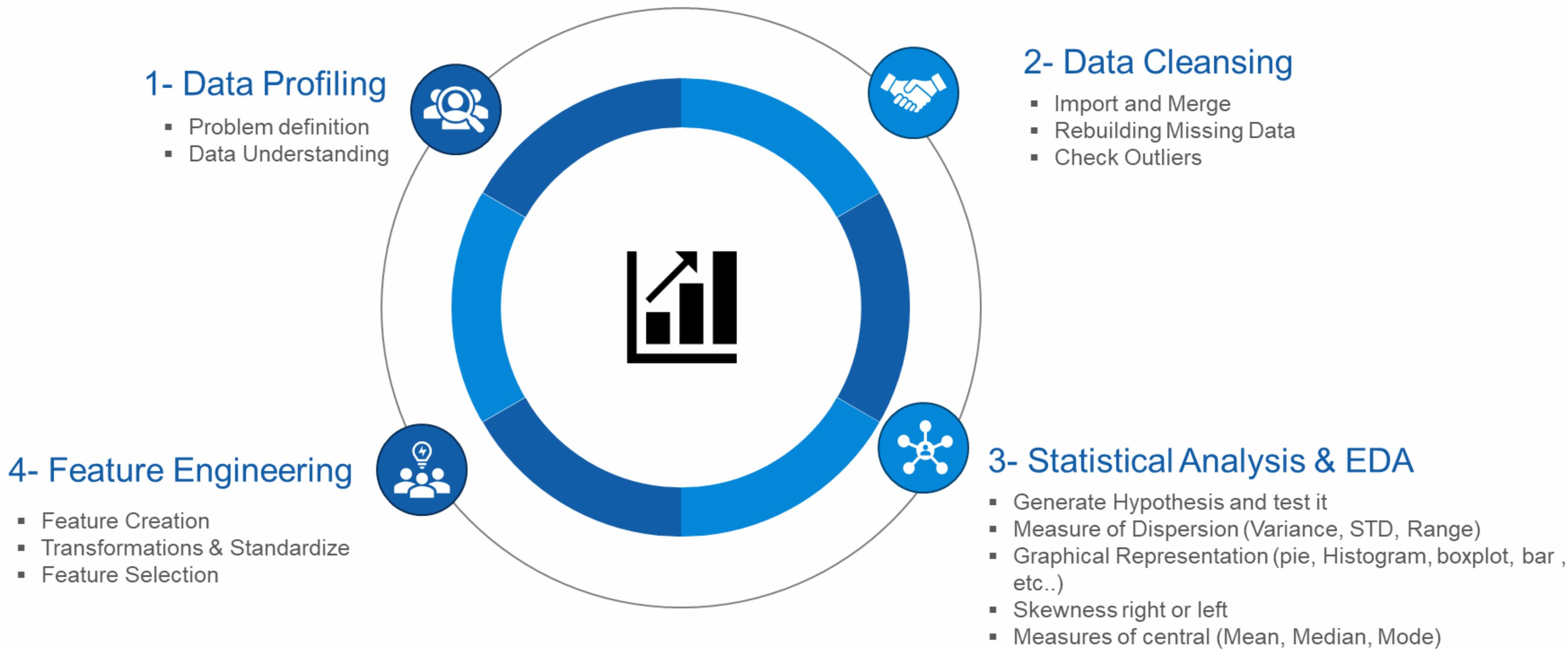
Competition
& Online
Competition

Supply Chain
Challenges

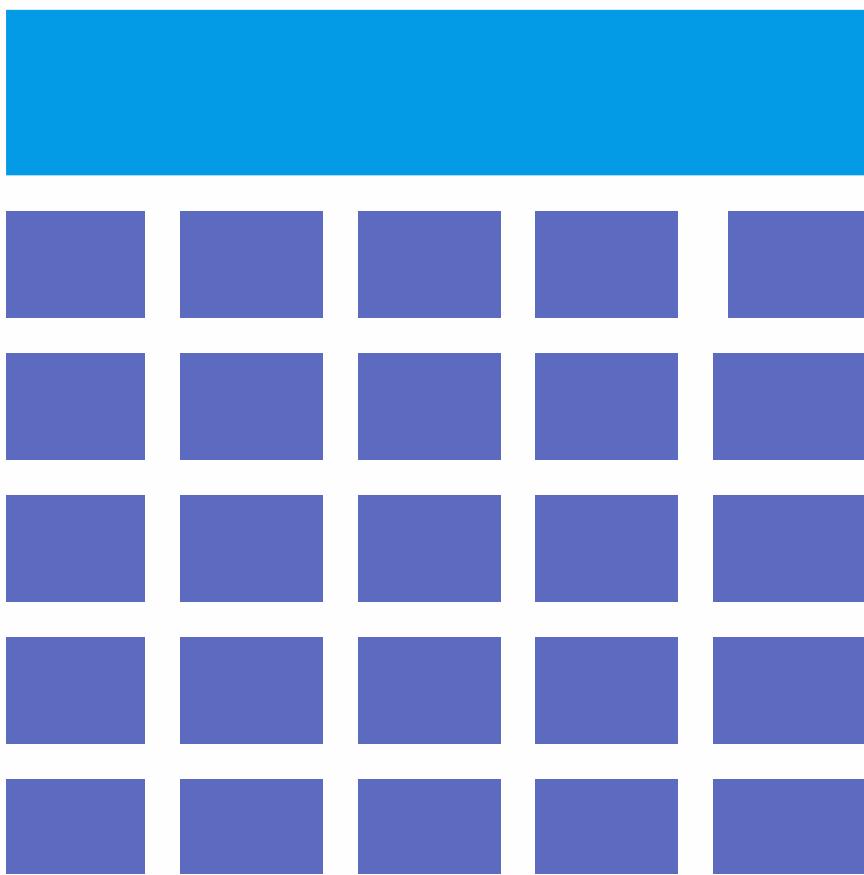
Economic
Factors

Changing
Consumer
Preferences

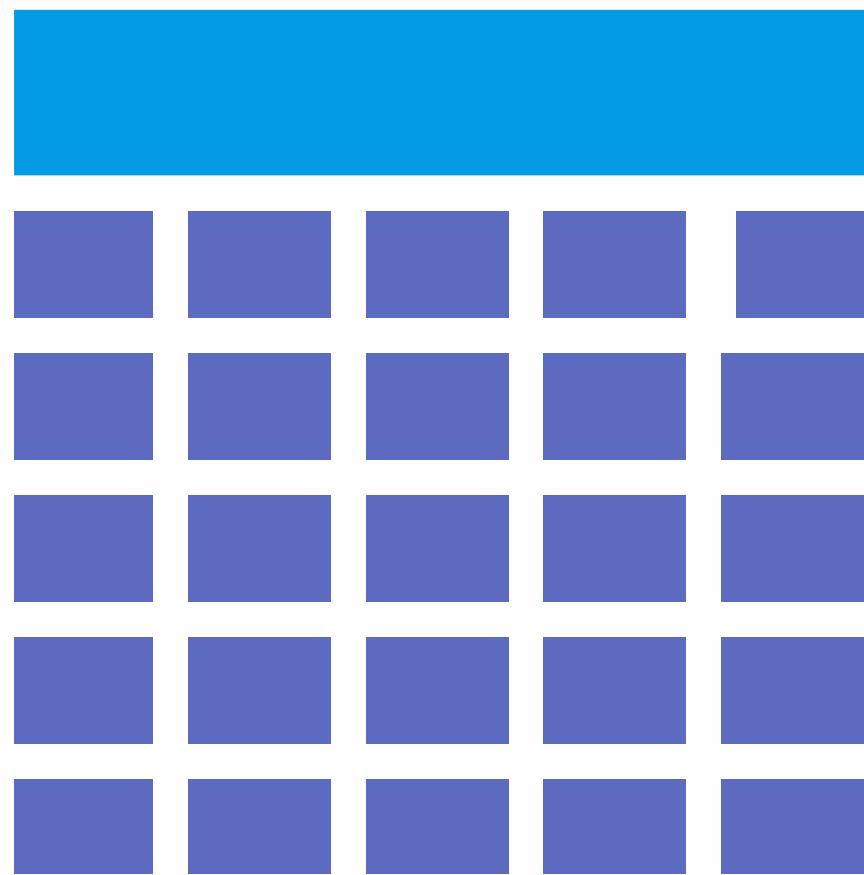
PROBLEM STRUCTURE



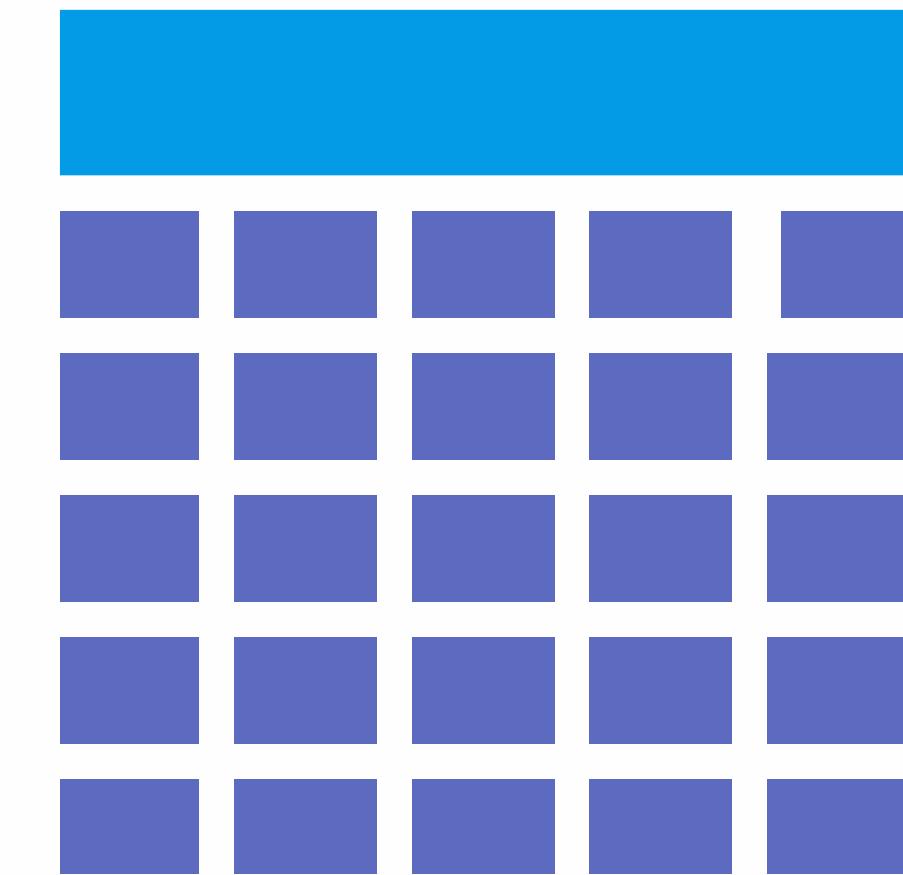
TASK IDENTIFICATION



STORES



FEATURES



TRAIN

STORES

(45, 3)



STORES



STORE TYPE



STORE SIZE

STORES - STORE



01

Data Type

- inr64

02

Unique

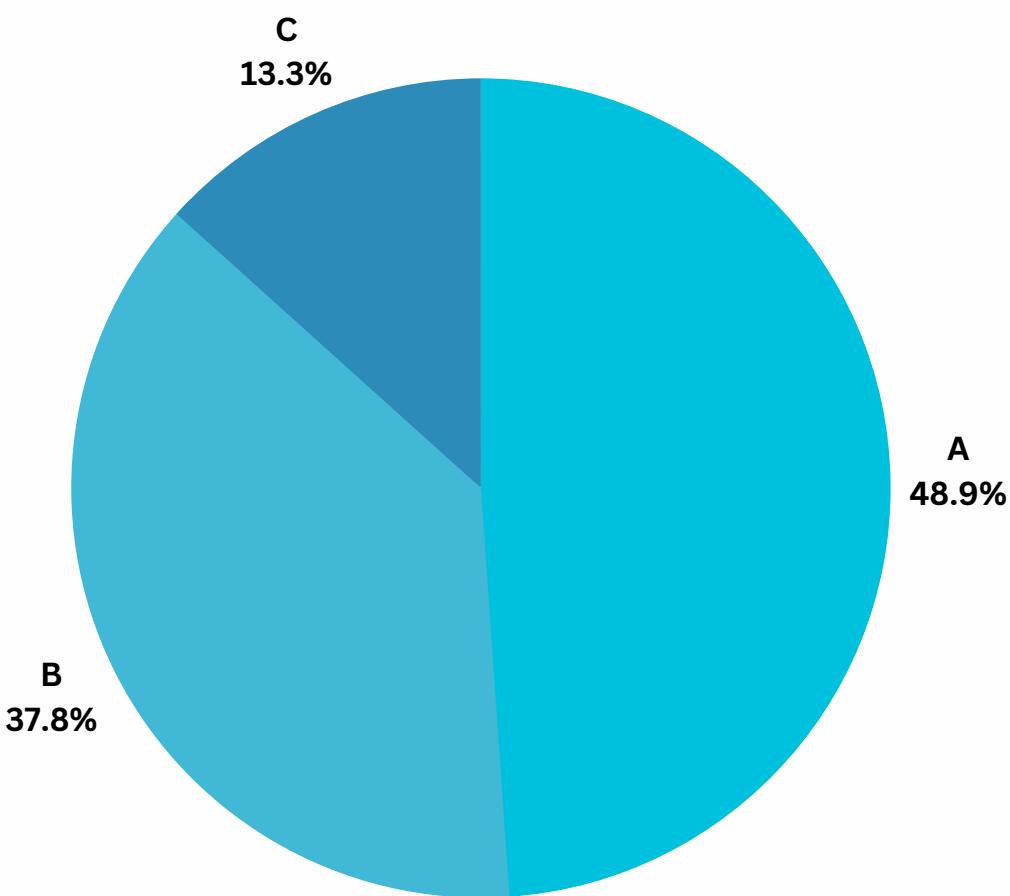
- 45 (1-45)

03

Missing & Outliers

- Missing : 0
- Outliers: 0 “ Categorical Data “

STORES - TYPE



01

Data Type

02

Unique

03

Missing & Outliers

- Object

- 3 (A, B ,C)

- Missing : 0.00%
- Outliers : 0.00% “Categorical Data”

STORES - TYPE

01

Supercenters

02

Neighborhood
markets

03

Discount
stores

STORES - SIZE



01

Data Type

- int64

02

Unique

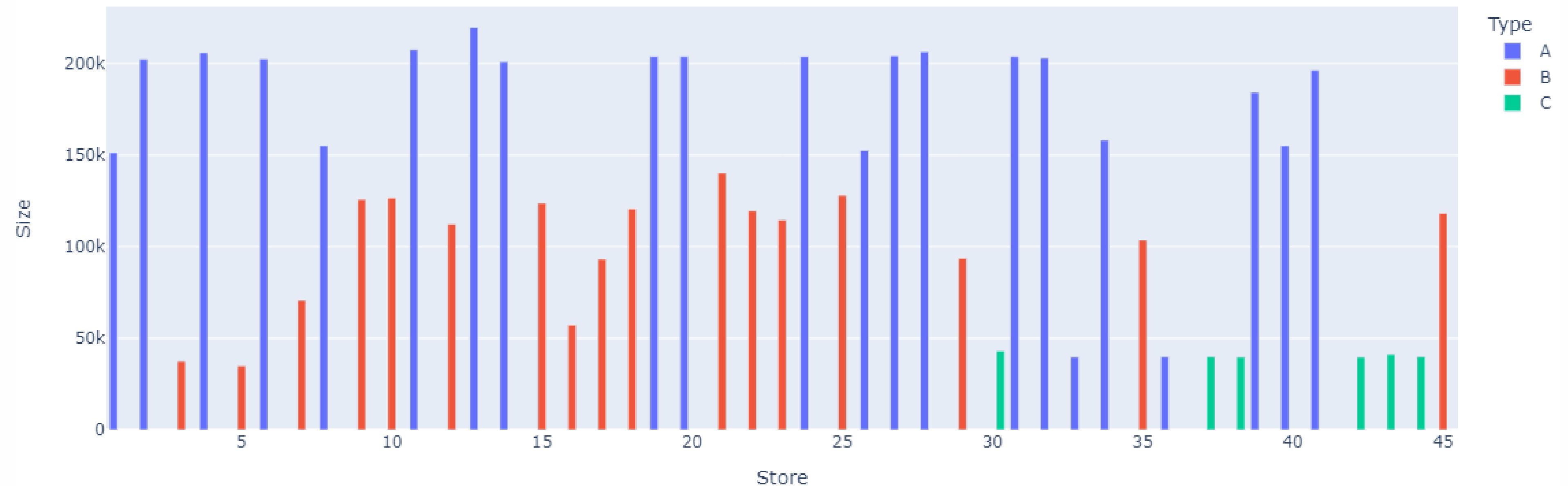
- Continuous Data

03

Missing & Outliers

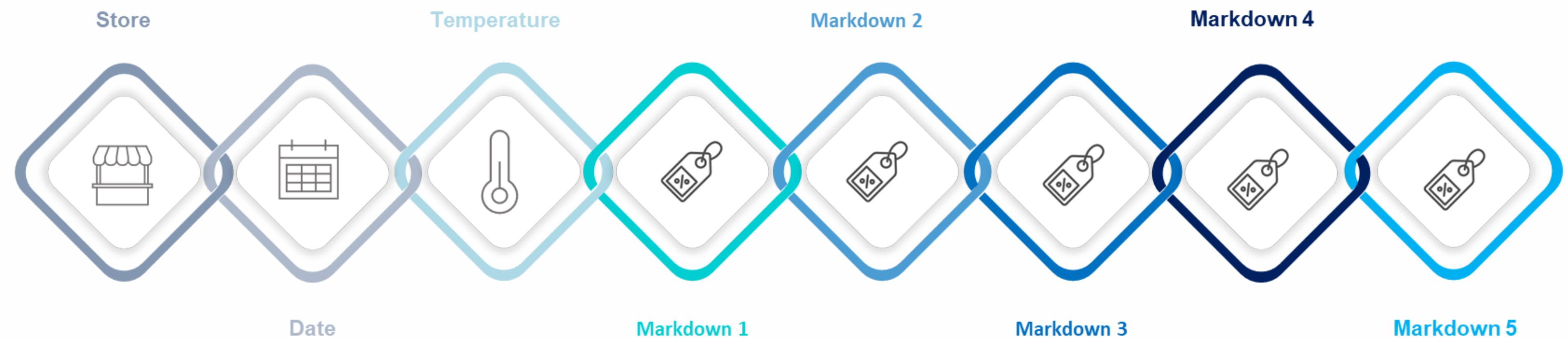
- Missing : 0
- Outliers: 0

STORE SIZE



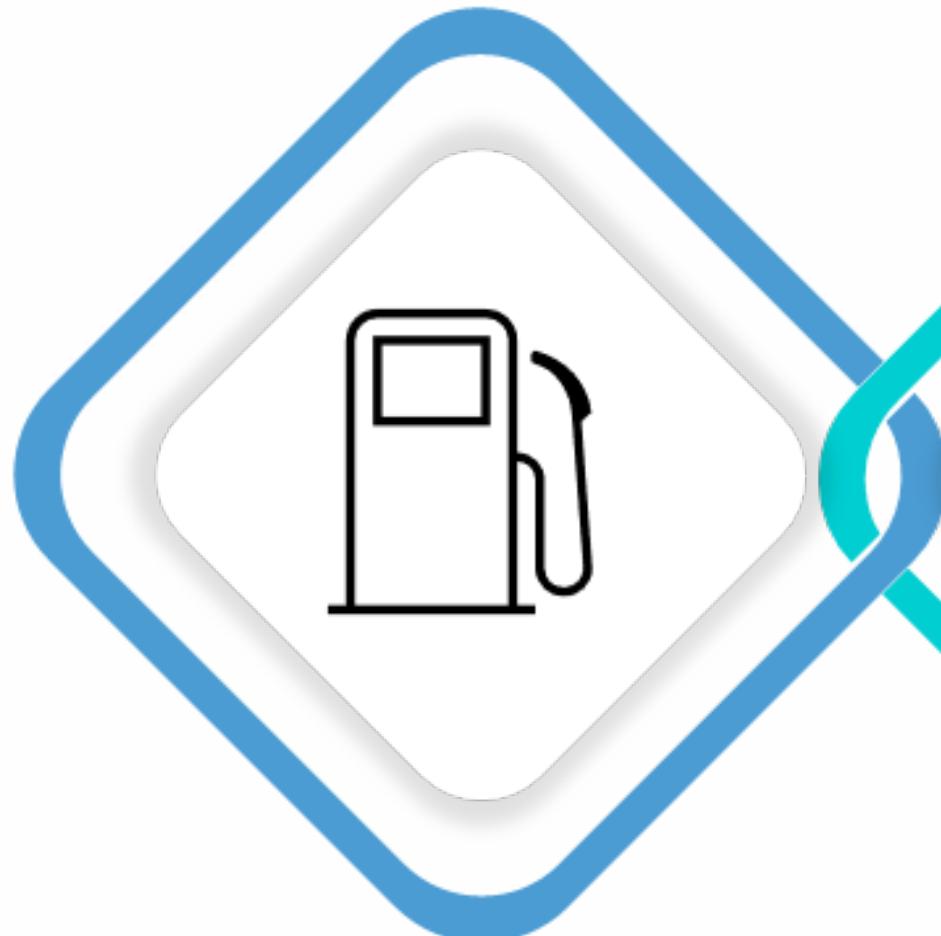
FEATURES

(8190,12)

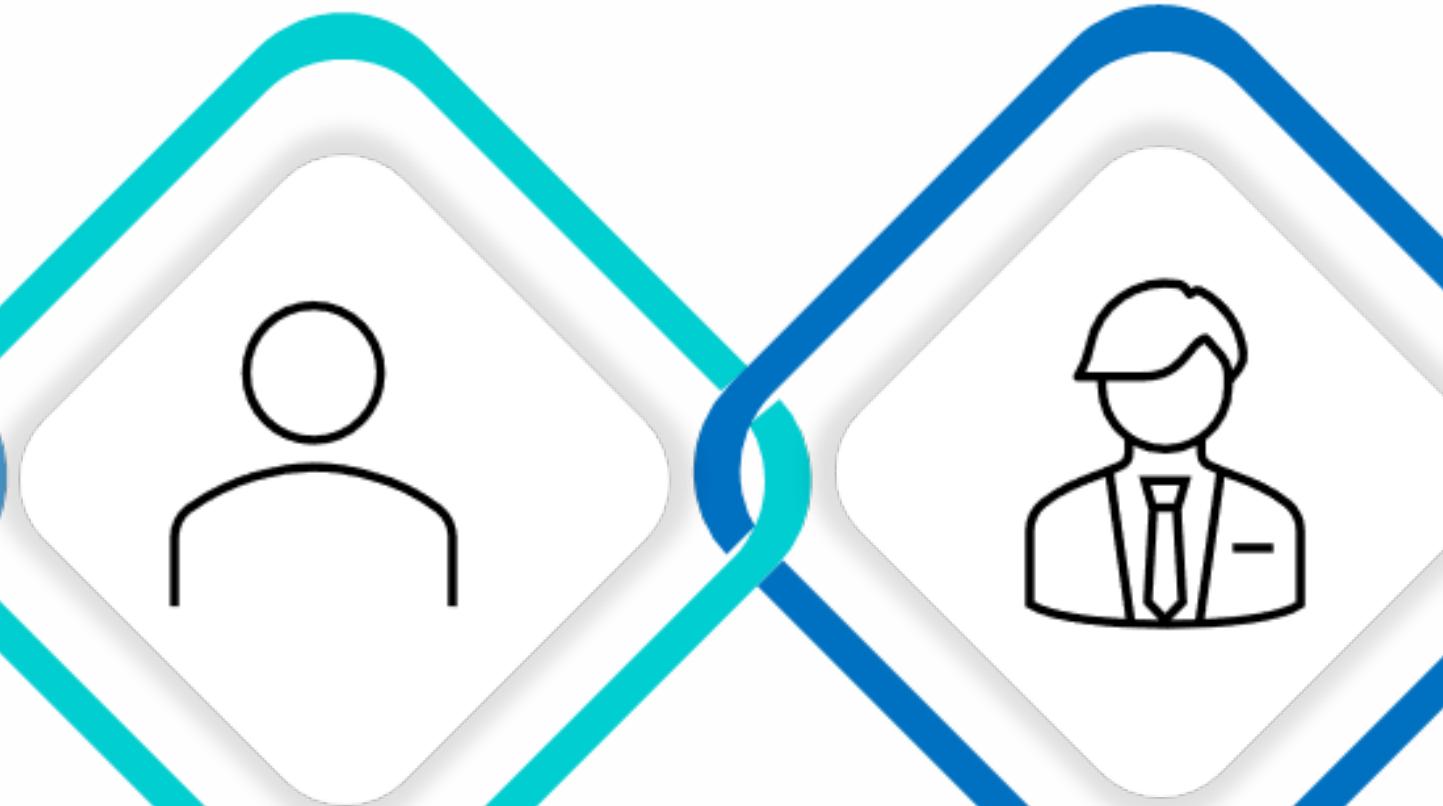


FEATURES

Fuel Price



Unemployment



CPI



Is Holiday

FEATURES - DATE



01

Data Type

- Object

02

Unique

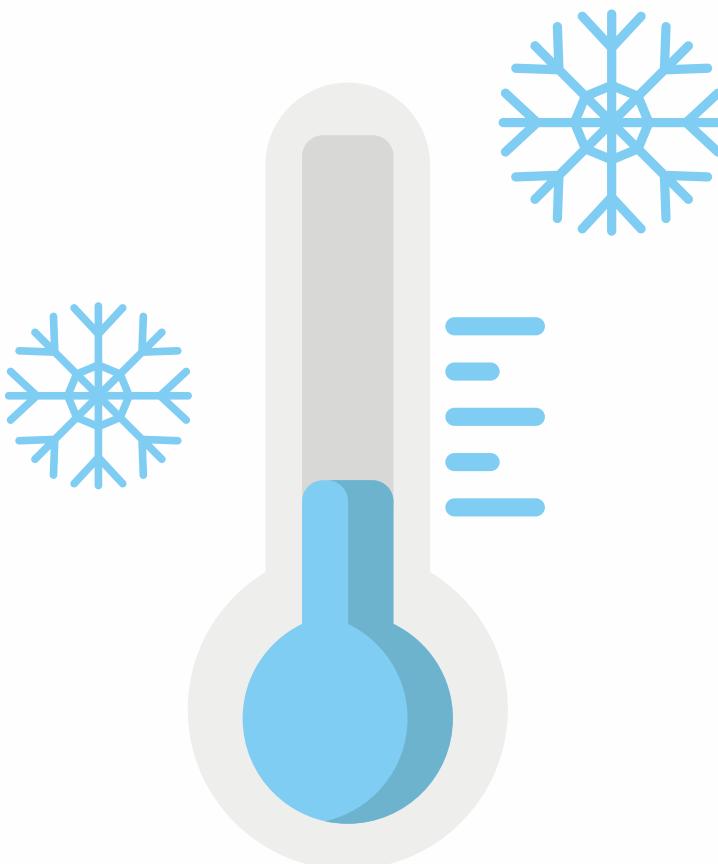
- Date
- From 05-02-2010 to 26-10-2012

03

Missing & Outliers

- Missing : 0.00%
- Outliers : 0.00% "Date Data"

FEATURES - TEMPERATURE



01

Data Type

- Float64

02

Unique

- Continuous

03

Missing & Outliers

- Missing : 0
- Outliers: 0.02

FEATURES - MARKDOWN 1



01

Data Type

- Float64

02

Unique

- Continuous

03

Missing & Outliers

- Missing : 50.76%
- Outliers: 35.74%

FEATURES - MARKDOWN 2



01

Data Type

- Float64

02

Unique

- Continuous

03

Missing & Outliers

- Missing : 50.76%
- Outliers: 64.33%

FEATURES - MARKDOWN 3



01

Data Type

- Float64

02

Unique

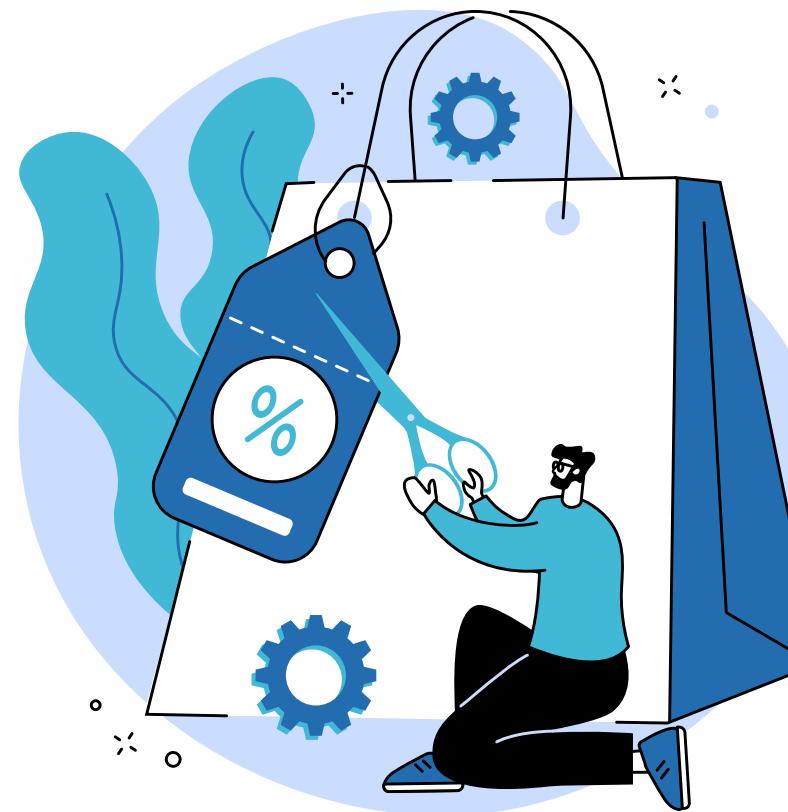
- Continuous

03

Missing & Outliers

- Missing : 55.88%
- Outliers : 32.50%

FEATURES - MARKDOWN 4



01

Data Type

- Float64

02

Unique

- Continuous

03

Missing & Outliers

- Missing : 57.70%
- Outliers : 32.02%

FEATURES - MARKDOWN 5



01

Data Type

- Float64

02

Unique

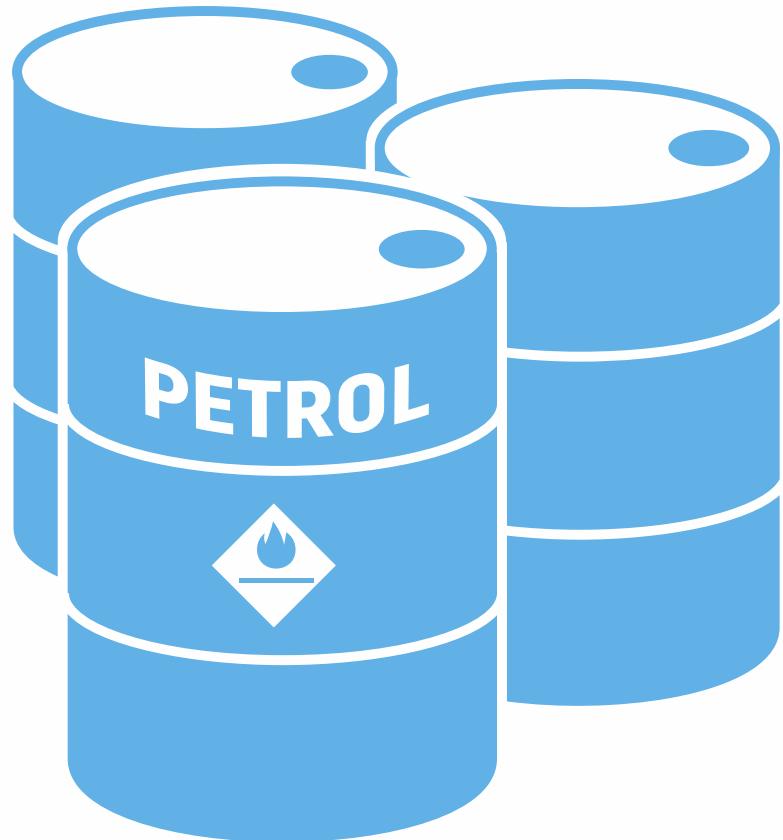
- Continuous

03

Missing & Outliers

- Missing : 50.54%
- Outliers : 35.92%

FEATURES - FUEL PRICE



01

Data Type

- Float64

02

Unique

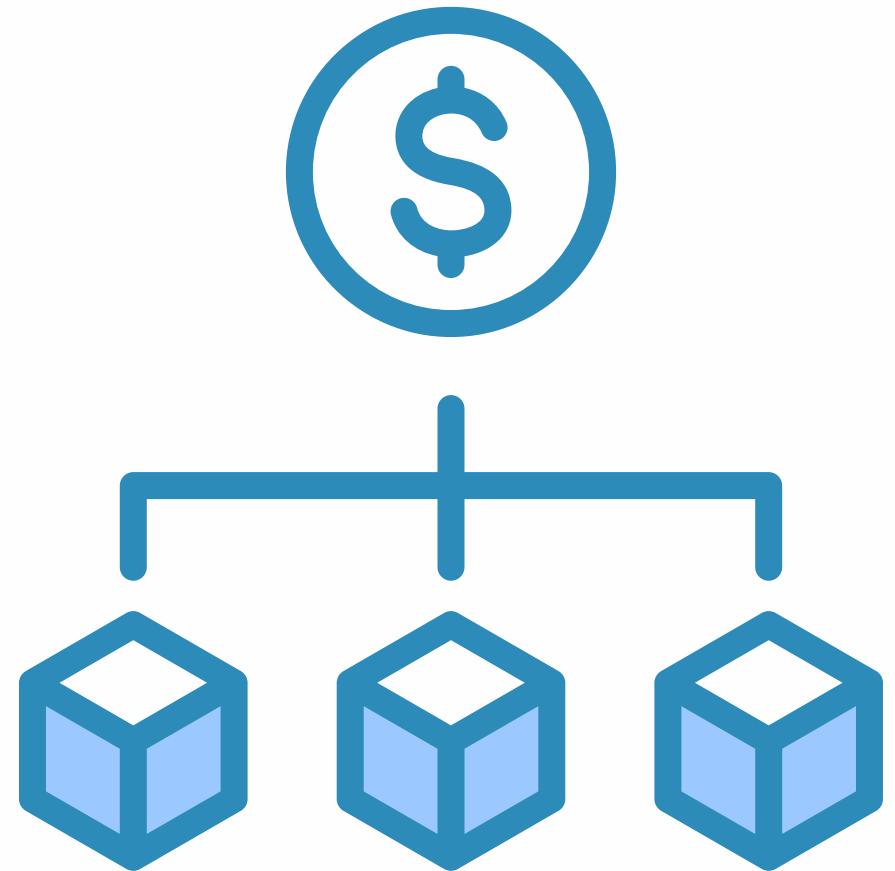
- Continuous

03

Missing & Outliers

- Missing : 0.00%
- Outliers : 0.00%

FEATURES - CPI



01

Data Type

02

Unique

03

Missing & Outliers

- Float64

- Continuous

- Missing : 7.14%
- Outliers : 0.00%

FEATURES - UNEMPLOYMENT



01

Data Type

- Float64

02

Unique

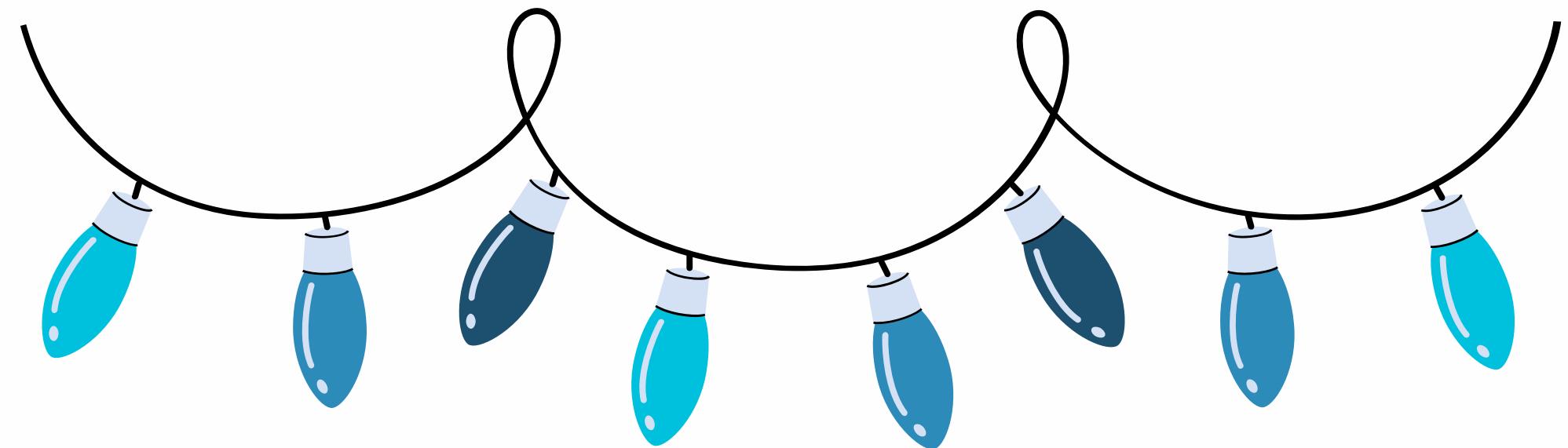
- Continuous

03

Missing & Outliers

- Missing : 7.14%
- Outliers : 7.62%

FEATURES - IS HOLIDAY



01

Data Type

- Boolean

02

Unique

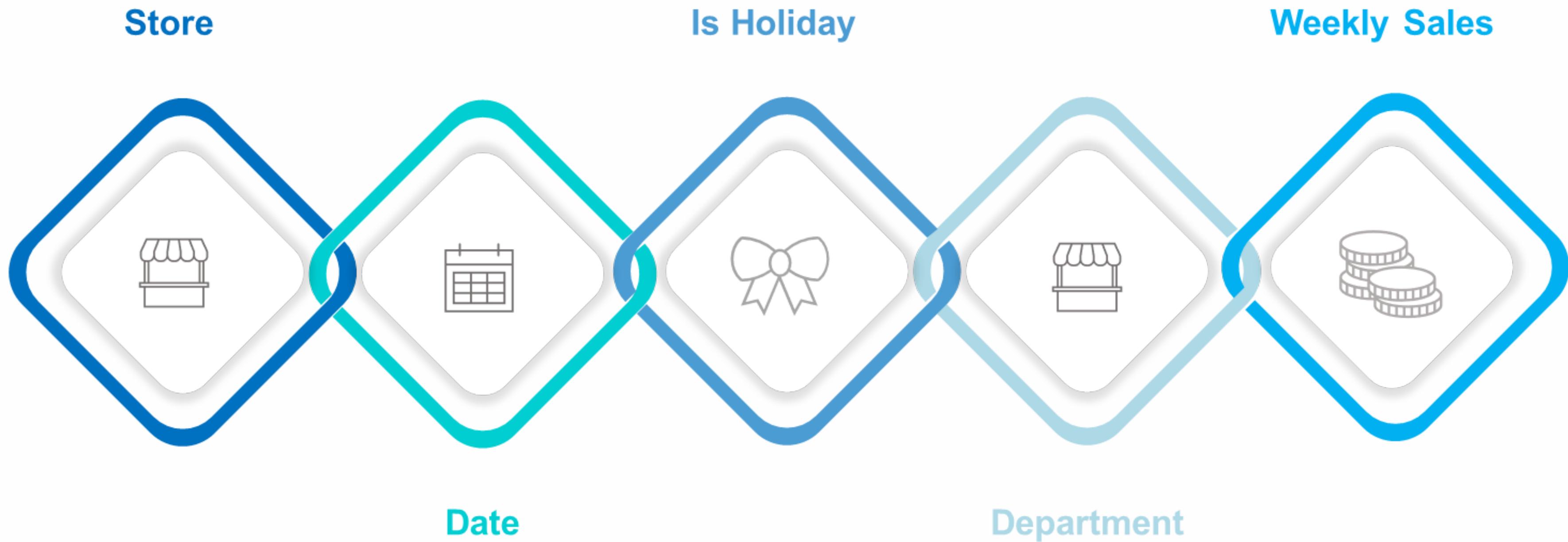
- 2 (True, False)

03

Missing & Outliers

- Missing : 0.00%
- Outliers : Boolean data

TRAIN [421570, 5]



TRAIN - DEPARTMENT



01

Data Type

- Int64

02

Unique

- 99 unique (1-99)

03

Missing & Outliers

- Missing : 0.00%
- Outliers : Categorical data

TRAIN - WEEKLY SALES



01

Data Type

02

Unique

03

Missing & Outliers

- Float64

- Continuous

- Missing : 0.00%
- Outliers : 8.43%

DATA CLEANSING

MERGE THE TABLES



DEELING WITH MISSING

01

Median

- Markdown 1
- Markdown 2
- Markdown 3
- Markdown 4
- Markdown 5

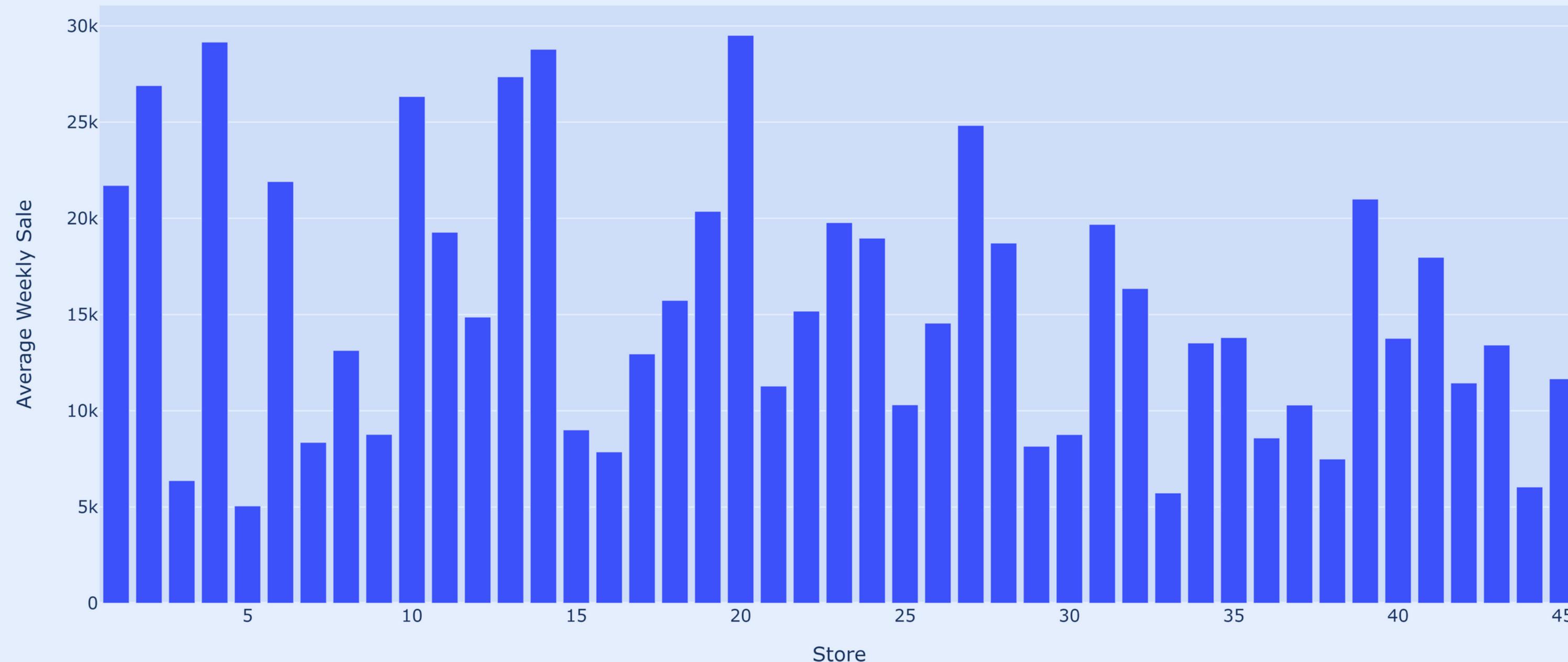
02

Last Recorded

- CPI
- Unemployment

HISTORICAL SALES

Average Weekly Sale by Store



HISTORICAL SALES

Average Sales for Store by Month

	2	4	6	8	10	12						
Store	9927.0	11505.7	11302.7	11678.8	11697.1	12090.7	11284.6	10936.5	10679.8	10958.8	12980.6	15645.2
45	6081.9	6051.6	5902.3	5835.9	6126.1	6186.7	6089.8	6127.3	6030.3	5989.4	6064.7	6020.4
	13365.1	13615.0	13410.3	13561.2	13873.5	13670.4	13091.3	13489.8	13719.4	13305.0	13071.7	12553.2
	12068.8	11916.7	11552.5	11259.6	11532.7	11311.8	11151.9	11201.4	11317.4	11510.7	12007.2	10929.6
40	15680.2	16246.2	16083.9	17047.2	17914.0	18639.0	18283.3	19630.4	17388.0	17636.7	19845.0	22041.0
	11564.2	13331.8	12917.4	13017.5	13707.7	14442.3	14476.5	14115.6	13250.9	13524.6	14272.5	16471.8
	18843.1	19158.9	19727.7	20341.8	20333.6	21484.2	20656.9	22469.6	20204.5	20473.9	23318.2	26046.0
	7747.5	7465.1	7527.7	7433.3	7610.7	7411.4	7306.2	7482.4	7602.8	7570.7	7679.9	7212.1
	10730.7	10667.8	10049.3	10318.6	10309.4	10012.8	9978.5	10108.3	10182.6	10460.0	10689.1	10492.4
	8063.1	9535.4	8906.3	9264.1	8922.7	8832.3	8767.9	8438.8	7943.8	7899.8	8267.5	7831.3
35	9710.8	14020.3	13552.5	13838.0	13935.9	14783.6	14434.5	13888.7	12483.1	12326.1	15909.8	16618.8
	12505.1	13746.8	13177.0	13369.0	13534.0	13555.2	12851.0	13319.9	12943.7	13222.1	14696.6	16029.5
	5577.3	6044.4	5858.2	6090.2	6265.0	5867.2	5775.2	5484.9	5511.4	5452.4	5491.8	5112.5
	14979.1	15620.0	15391.4	15717.4	16245.9	16569.9	16184.1	16930.0	15602.2	16078.5	17907.2	19852.5
	18890.3	20561.5	19486.6	19326.1	19370.8	19870.3	19052.7	19696.1	18608.6	18853.8	21421.0	22068.0
30	8960.9	9017.8	8667.8	8703.9	8989.3	8798.5	8587.9	8603.9	8547.2	8992.3	8899.3	8543.6
	6583.4	7903.2	7791.4	8046.8	8078.9	8596.1	8007.6	7905.2	7382.4	7685.5	9530.8	10912.5
	17018.5	21033.0	19403.7	18309.6	18039.5	18523.7	17938.0	18007.3	17465.5	17831.8	21175.0	20742.6
	20975.2	23359.3	22810.5	24863.0	25049.3	26308.1	25984.0	25323.5	23462.9	23375.8	27139.6	29515.8
25	12467.2	14020.8	13458.7	13553.4	14652.6	15102.2	15492.2	15374.8	14249.3	14252.8	15208.0	16740.5
	8640.8	9173.5	9477.6	10204.9	10458.1	10639.7	10283.8	10341.3	9548.1	9887.5	11842.5	13888.3
	16078.6	18277.1	17479.5	18008.9	18784.5	19775.1	20200.0	19474.2	18288.9	18508.5	20253.7	22652.3
	15590.9	18224.2	17880.5	18701.4	19393.3	21266.2	19961.4	21206.3	18544.8	18595.6	21791.5	26837.8
	12537.3	14405.1	14291.3	14912.3	14990.9	15873.9	15063.2	15122.1	13943.9	14685.5	17195.4	19938.2
	9530.8	11555.5	11148.8	11077.8	10793.7	11355.5	10977.1	11845.0	10023.5	10145.2	12849.3	14835.7
20	26149.6	30464.0	28236.0	28786.6	28786.7	29803.1	28857.3	28592.8	27855.9	28541.7	33017.5	36708.0
	17251.8	19942.6	19179.3	19882.3	20298.7	20359.5	20161.9	20535.3	20214.9	19536.7	22259.1	25406.4
	12761.0	15990.6	15183.0	15415.1	15707.9	16354.3	15408.3	15655.8	13931.8	15248.4	17662.5	20053.3
	12494.1	11936.8	11752.1	12901.7	12969.9	13386.1	13297.0	12155.0	14309.4	12705.0	13468.1	14321.9
	7536.4	7215.2	7257.7	6471.1	6818.4	8533.6	8991.5	8597.6	7876.9	7398.1	7947.4	9988.7
15	7028.1	8592.3	8164.9	8694.7	9085.9	9534.2	9122.9	8920.1	8323.6	8314.7	10519.3	12258.2
	24974.1	28781.5	28363.3	29467.9	29231.3	29298.5	27665.3	27131.2	26536.9	27085.5	32264.1	36197.6
	24036.0	26467.0	25925.9	26705.4	26793.3	28022.7	27242.2	27707.3	25803.1	26540.7	29940.4	34308.0
	12851.7	16115.7	15031.2	14810.2	14646.9	14791.2	13978.7	14737.4	13587.9	13769.8	16779.4	18094.2
	16896.2	20809.3	19254.1	19240.3	18366.1	19050.6	18501.4	19773.8	17746.5	18059.1	20770.8	23635.1
10	23393.4	28407.5	26271.4	26271.6	25278.8	25009.8	25088.5	26132.8	23494.7	23930.2	29978.5	35134.3
	7678.3	8901.0	8495.4	8776.8	8846.2	9038.1	8267.7	8680.4	8239.0	8468.7	9791.6	10489.3
	11901.3	13987.4	12754.3	12956.2	12938.4	13119.8	12401.4	12987.3	12416.7	12711.1	14332.5	15766.7
	8536.6	7938.4	8200.7	6963.6	6496.1	8735.5	9801.1	9420.7	7899.8	7240.4	8953.1	10785.7
	18540.1	21303.8	21800.2	21773.6	21552.8	23395.8	22919.0	21631.7	19652.8	19701.7	23813.1	27462.3
5	4504.8	4983.2	4925.8	5134.7	5080.2	5198.3	4879.6	4978.0	4916.5	4754.2	5697.4	5776.8
	26745.4	30118.3	28227.6	27809.5	28395.5	28904.8	28063.2	29082.9	27925.5	28723.7	32405.4	35473.0
	5830.4	6810.2	6413.3	6216.8	6219.7	6442.1	6066.0	6229.0	6022.2	6076.5	6960.8	7487.9
	24072.8	28047.5	26383.6	26254.0	26512.7	27472.1	25954.5	26575.1	24860.1	25596.7	29388.0	33057.9
	19587.0	22497.5	21751.0	21601.5	21669.0	22112.9	20915.2	21656.1	20653.5	20711.2	23097.8	24805.5



12



FEATURE IMPACT

Size of Store

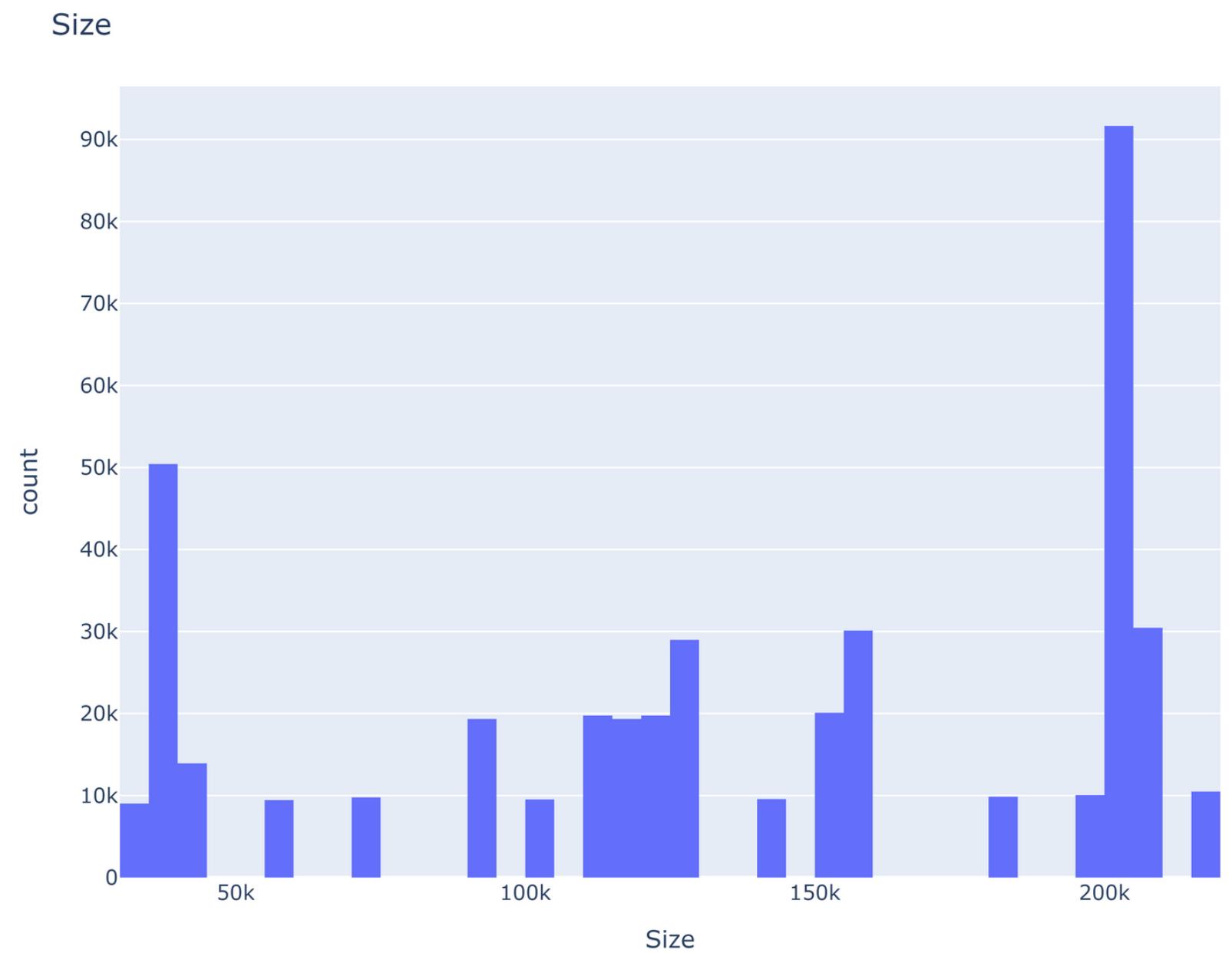
Box Plot - Outliers

Distribution

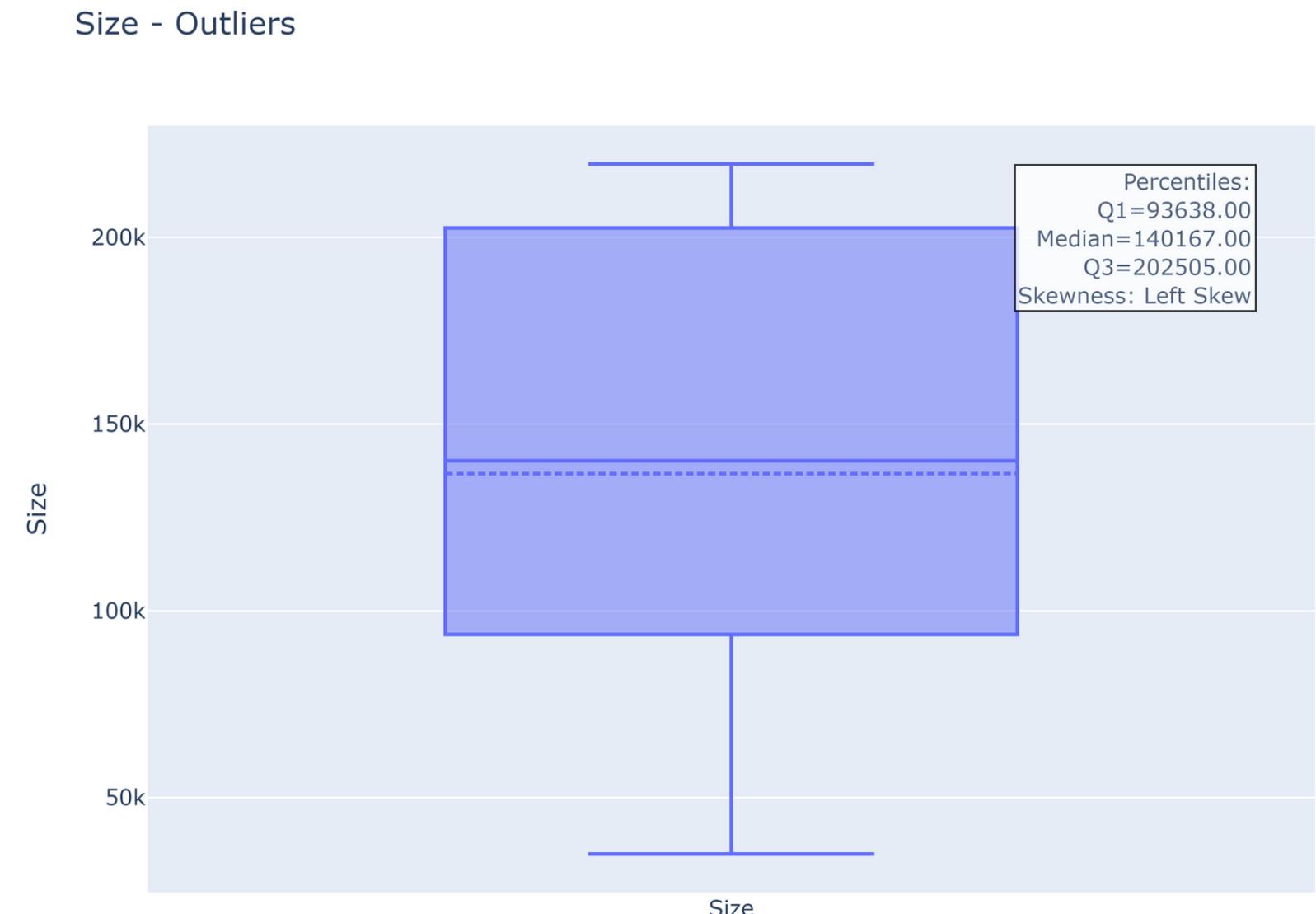
Hypothesis

Average Weekly
Sales

Distribution



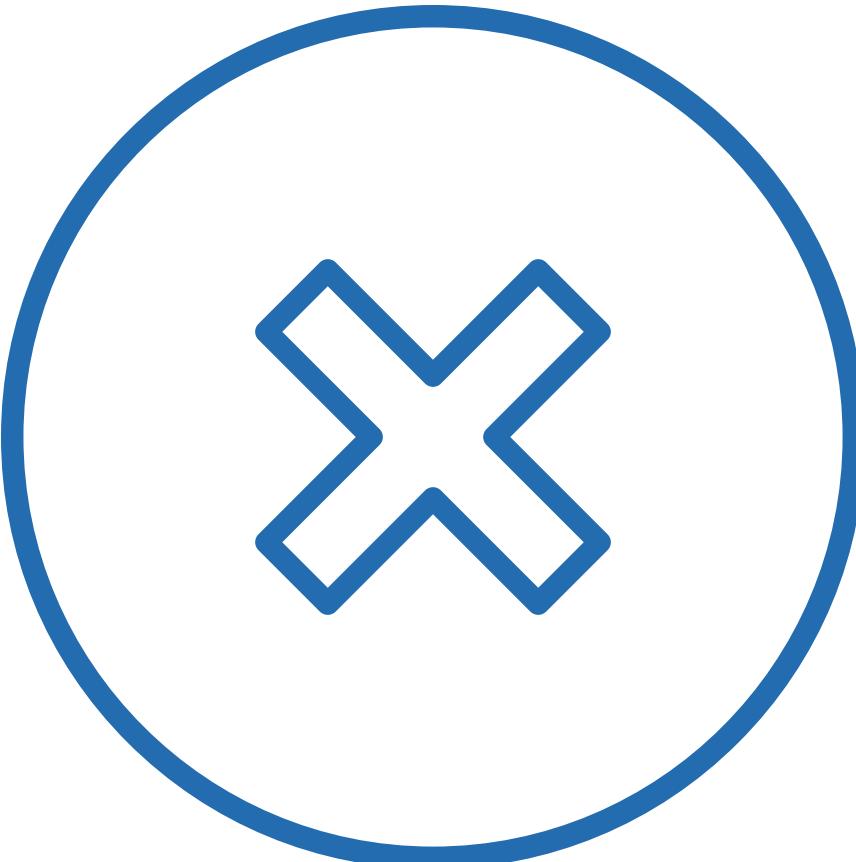
Box Plot - Outliers



Hypothesis 1

Null Hypothesis (H₀): Smaller or equal-sized stores have comparable weekly sales to larger stores.

Alternative Hypothesis (H_a): Smaller or equal-sized stores have noticeably different weekly sales compared to larger stores.



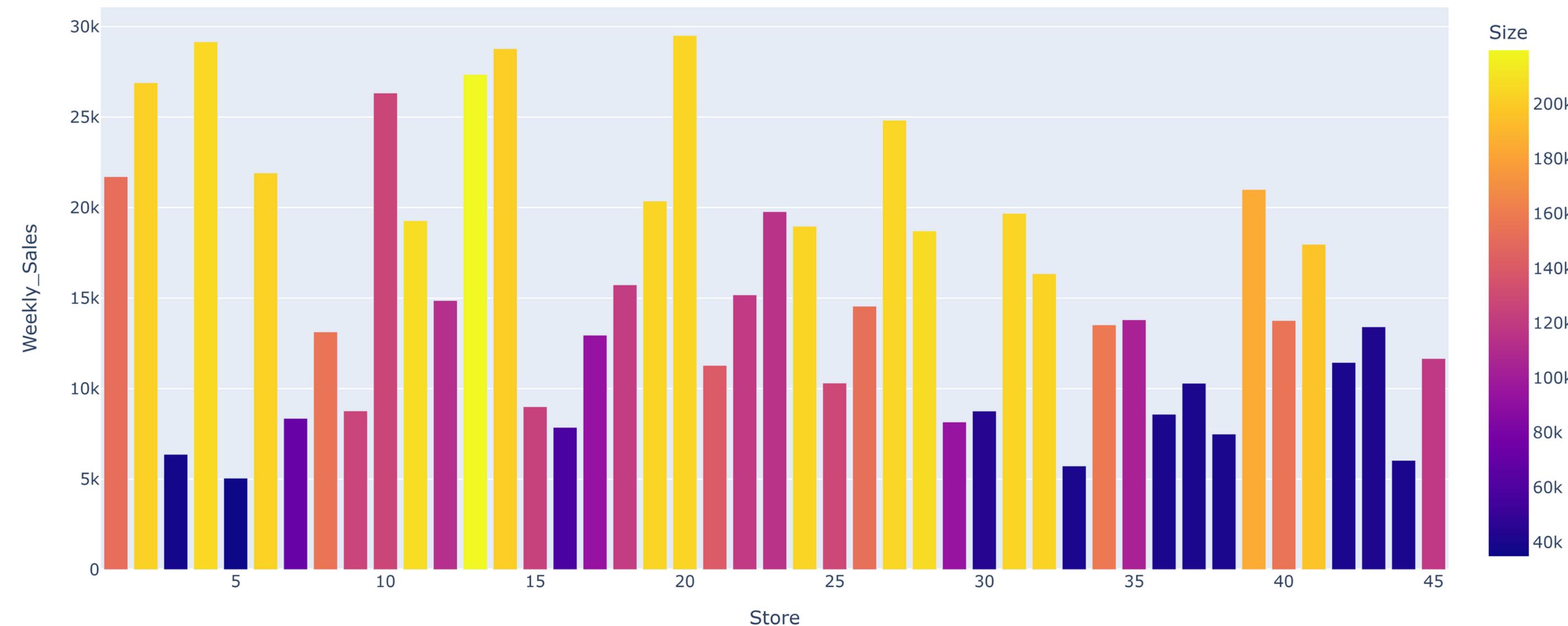
Reject null hypothesis: Stores smaller or equal in size to the mean store size have significantly different mean weekly sales compared to stores larger than the mean store size.

Significance level = 0.05

p_value = 0 z_statistic = -132.176

Size and Store vs Average Weekly Sales

Average Weekly Sales by Store and Size of Store



Usually the stores with greater size get a greater average weekly sales, also there are some stores with a medium size have a high weekly sales like store 10.

Type of Store

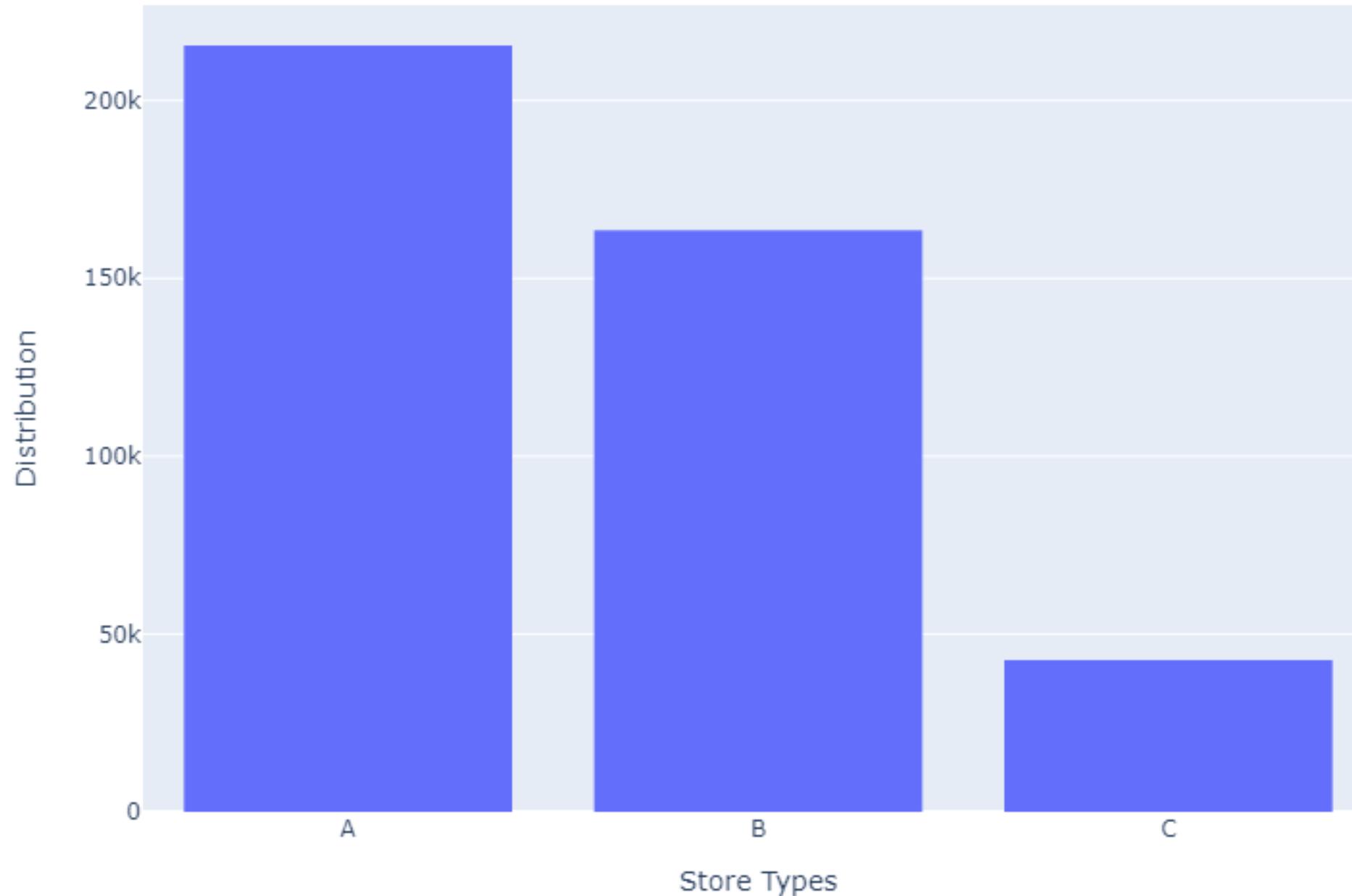
Distribution

Hypothesis

Average
Weekly Sales

Distribution

Distribution of Store Types

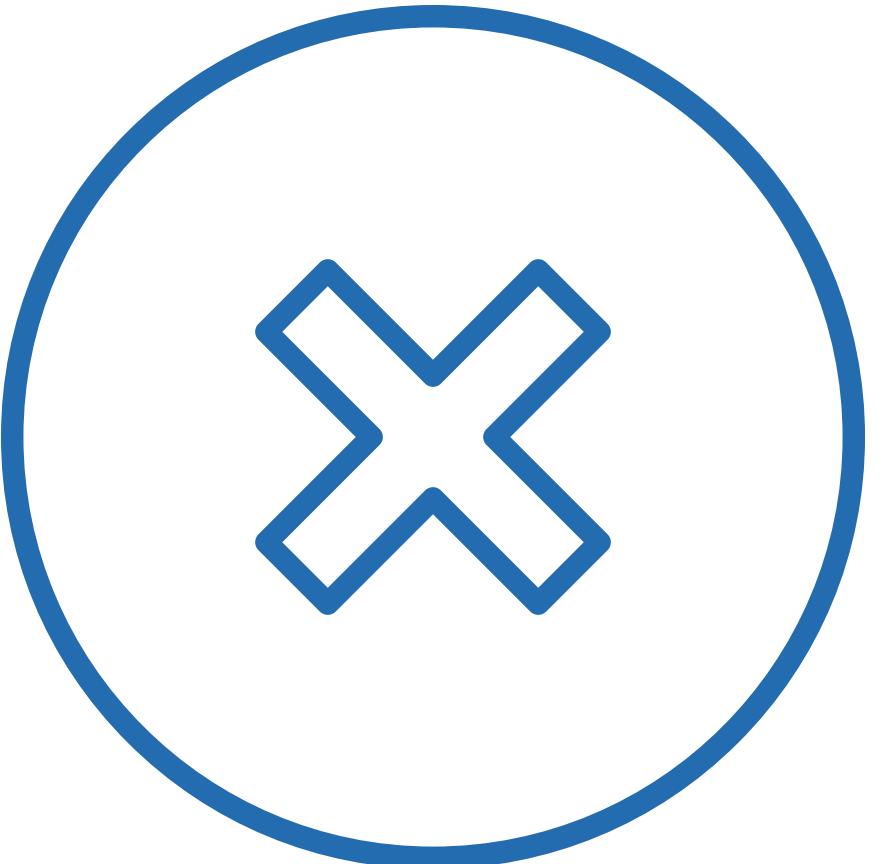


Store A : 22
Store B : 17
Store C : 6

Hypothesis 2

Null Hypothesis (H₀): There is no significant distinction in weekly sales between Type A and Type B stores.

Alternative Hypothesis (H_a): There is a significant contrast in weekly sales between Type A and Type B stores.



Reject null hypothesis: The average weekly sales for Type A stores differ significantly from the average weekly sales for Type B stores.

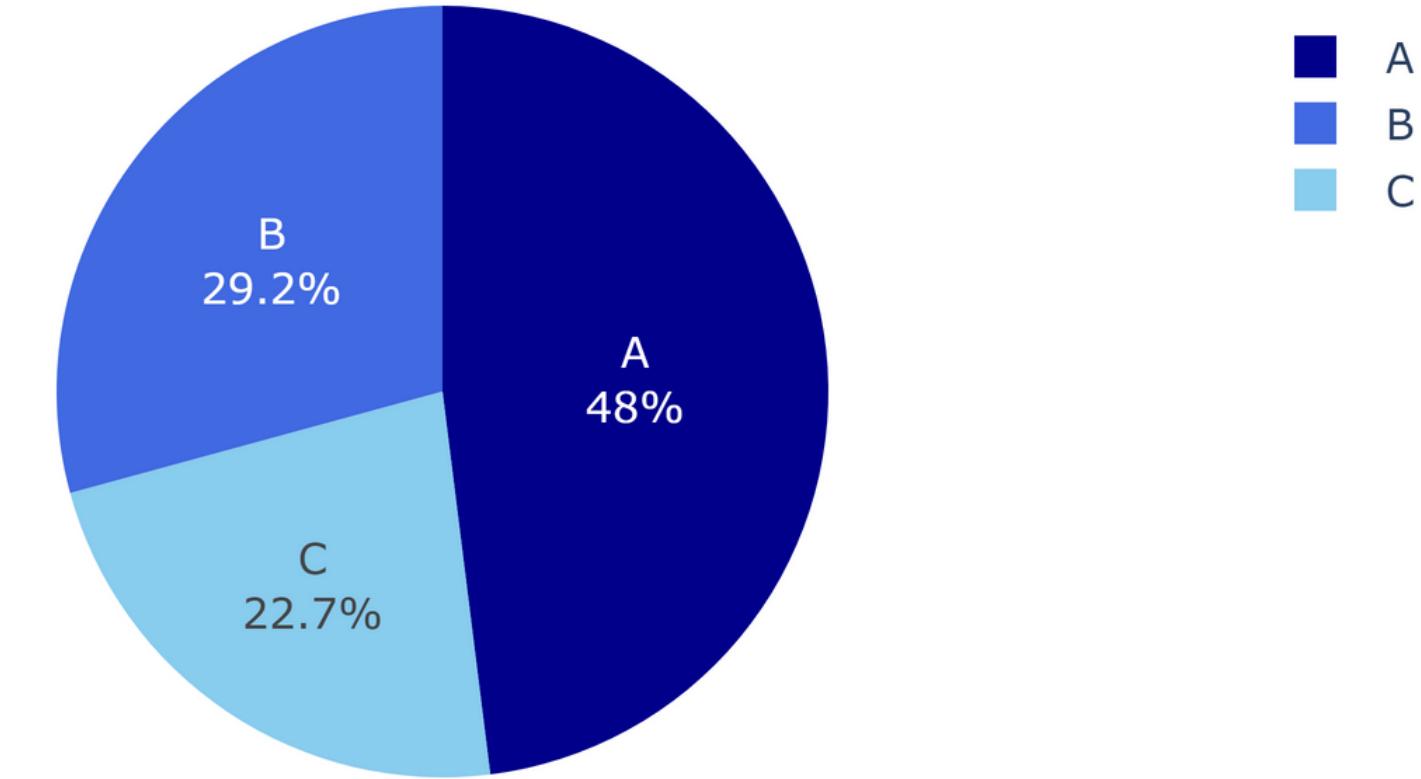
Significance level = 0.05

p_value = 0 z_statistic = 104.65

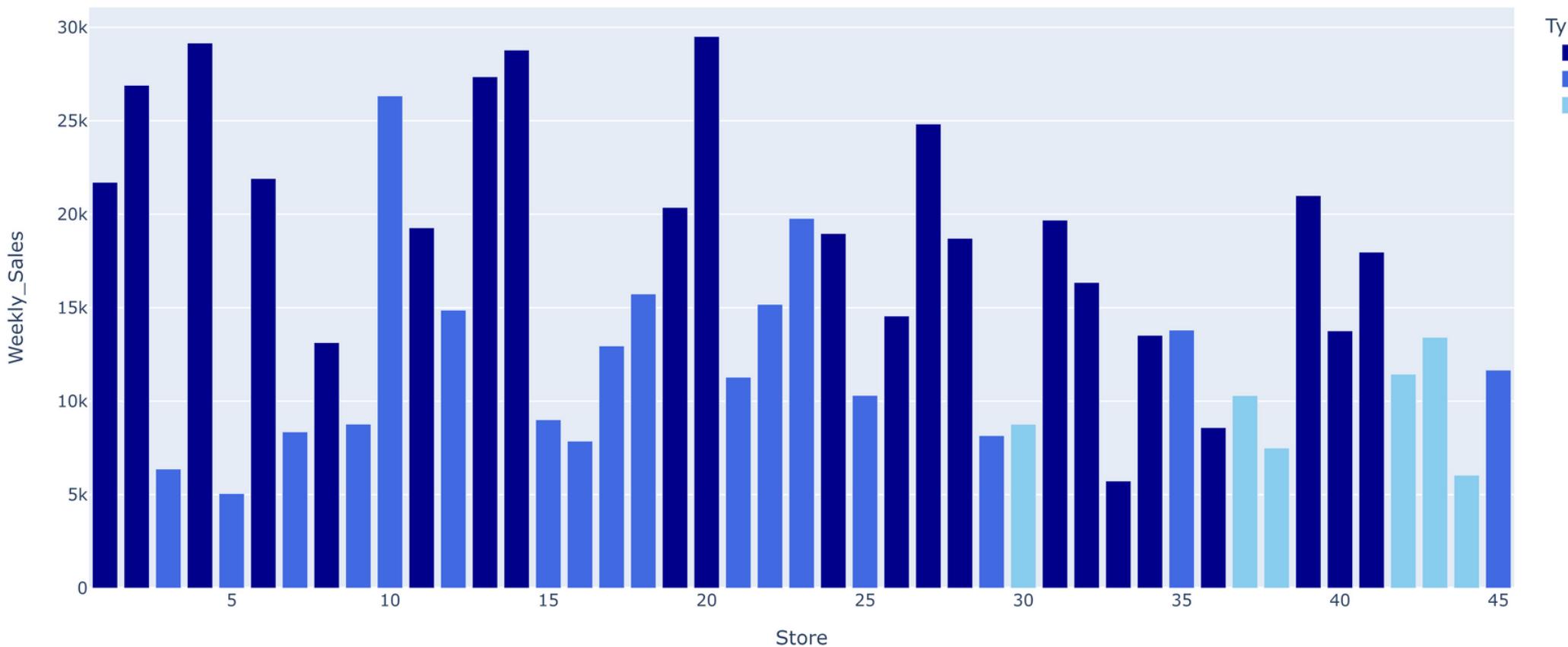
Average Weekly Sales by Type of Store

Type of Store

As we can see here, type A has the maximum average weekly sales in most stores; so that there is an actual difference between different weekly sales types.



Average Weekly Sales by Store and Type



Fuel Price

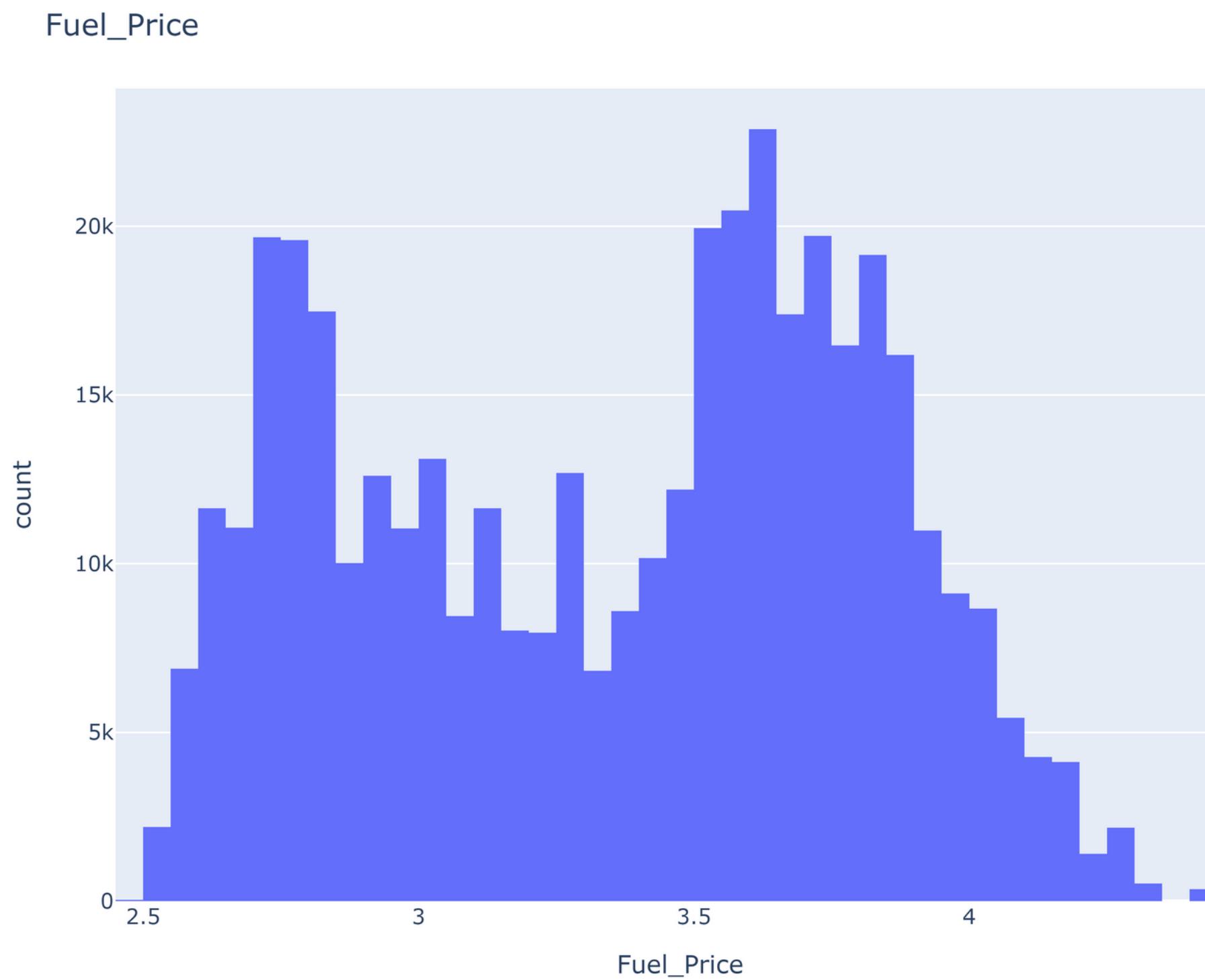
Distribution

Box Plot -
Outliers

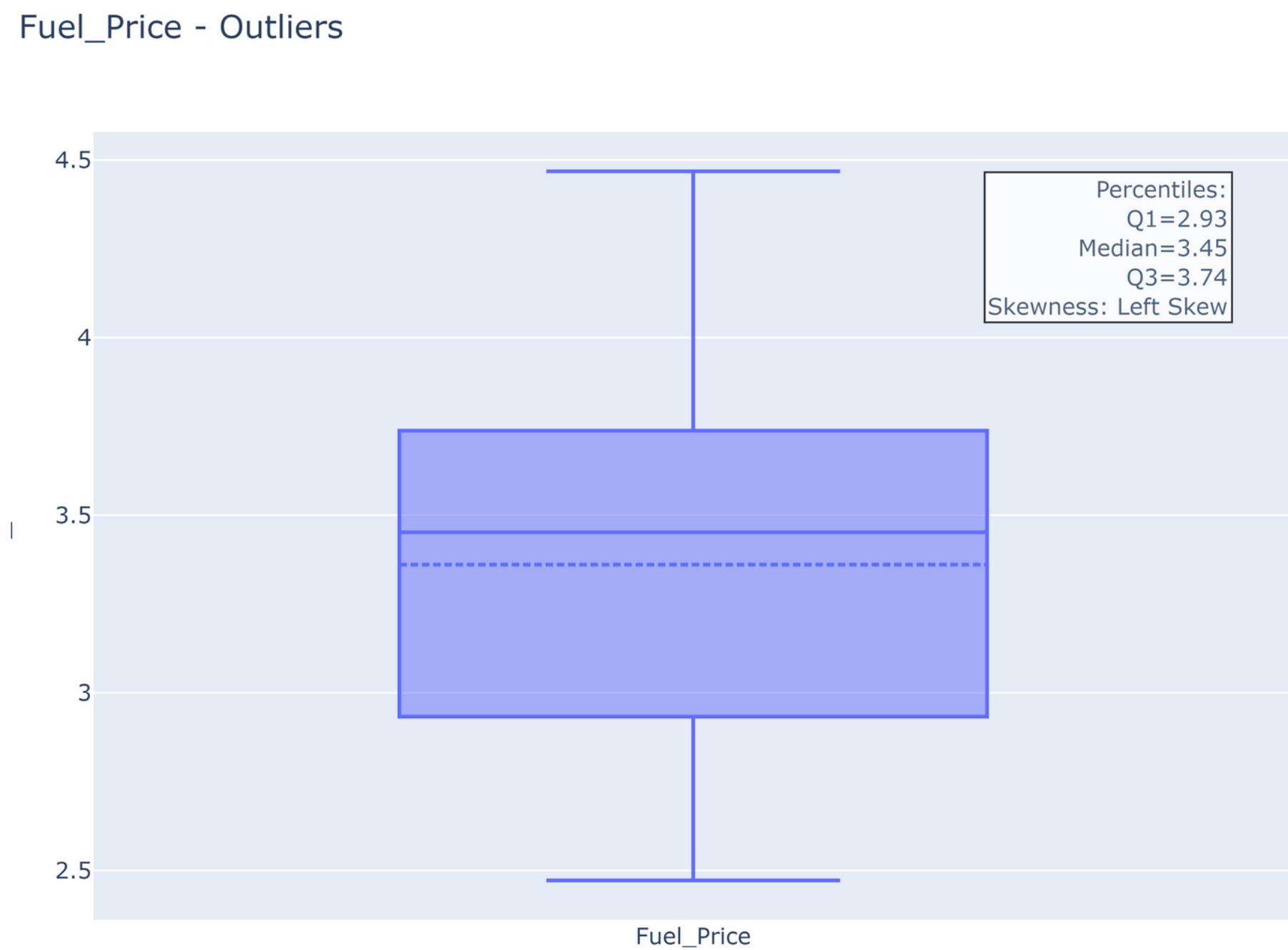
Average
Weekly
Sales

Hypothesis

Distribution



Box Plot - Outliers



Hypothesis 3

Null Hypothesis (H_0): There is no difference between the sales when fuel prices are above the average and when its below the average.

Alternative Hypothesis (H_a): There is a difference between the sales when fuel prices are above the average and when its below the average.



Reject null hypothesis: The average weekly sales during periods of high fuel prices differ from the average weekly sales during periods of low fuel prices.

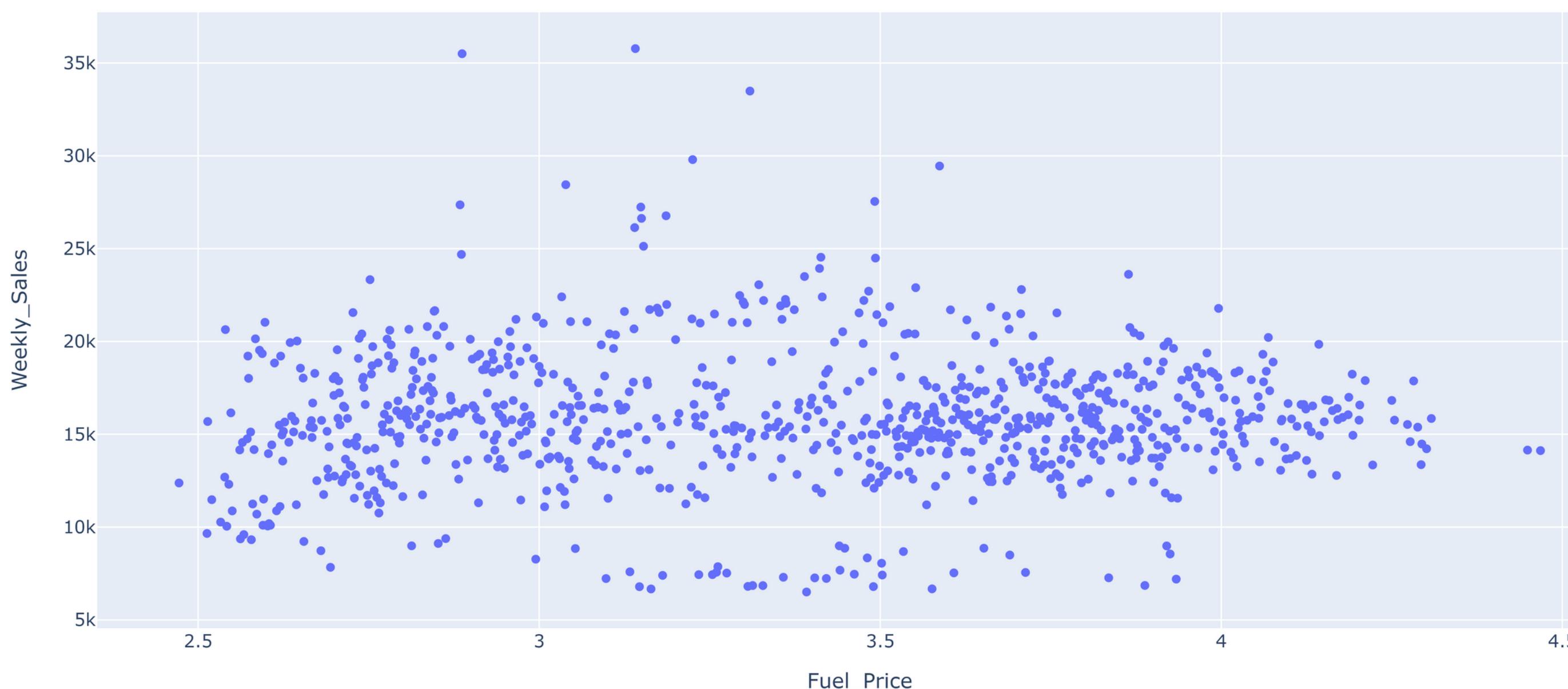
Significance level = 0.05

p_value = 0.01217

z_statistic = -2.5069

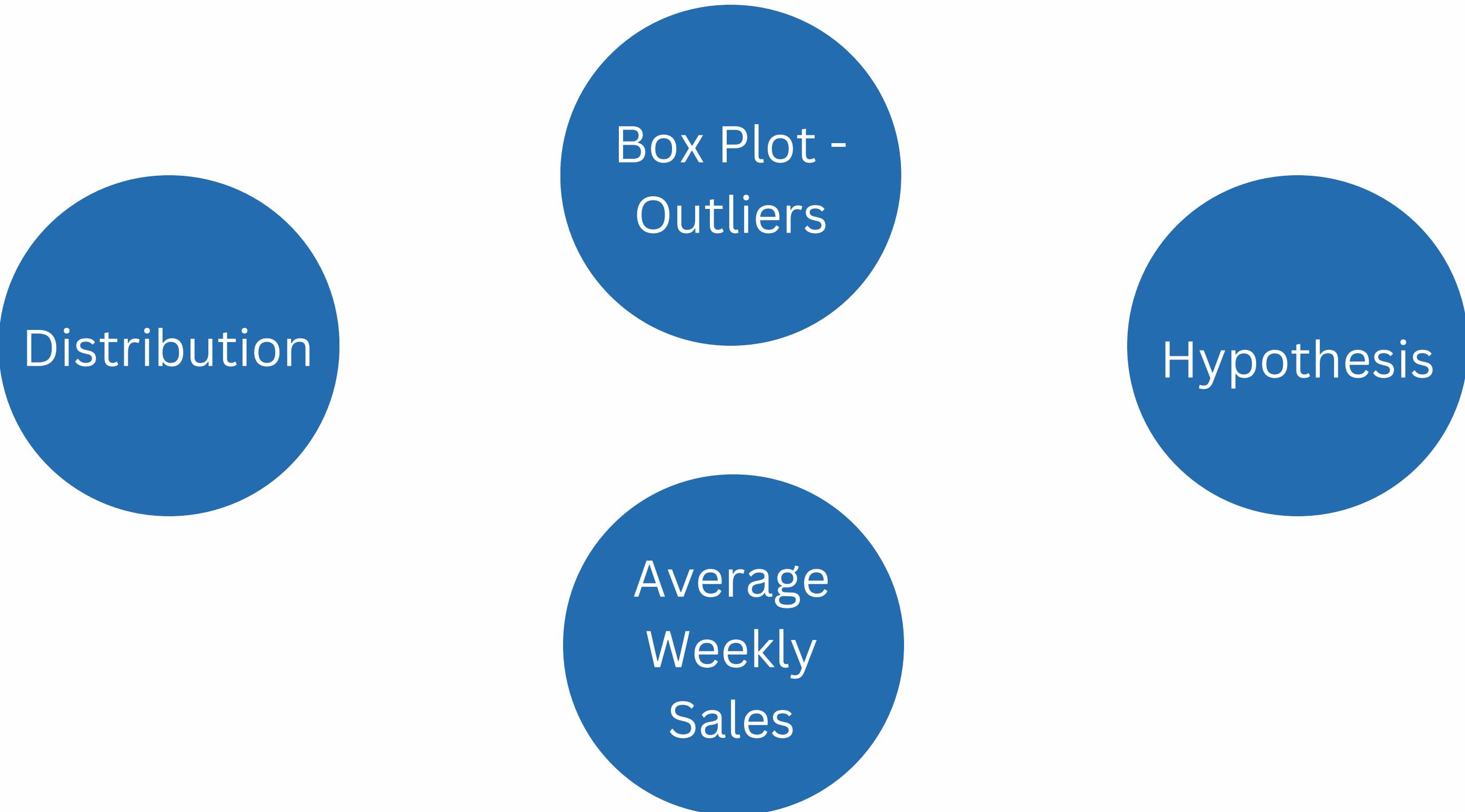
Fuel Price vs Weekly Sales

Average Weekly Sales by Fuel Price

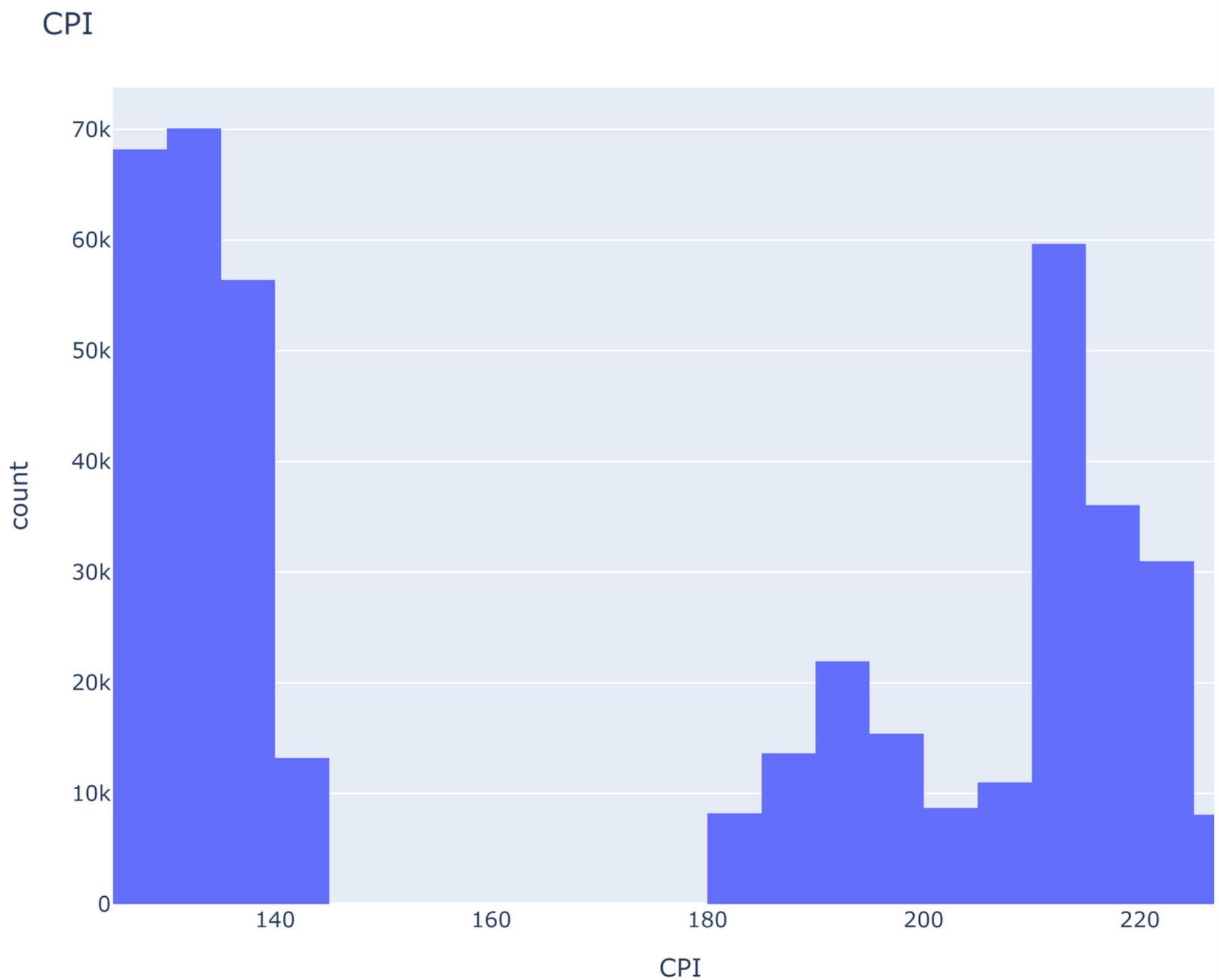


It is shown from this figure that fuel prices do really affect the weekly sales, when fuel prices are more than the average, the weekly sales are not increasing any more.

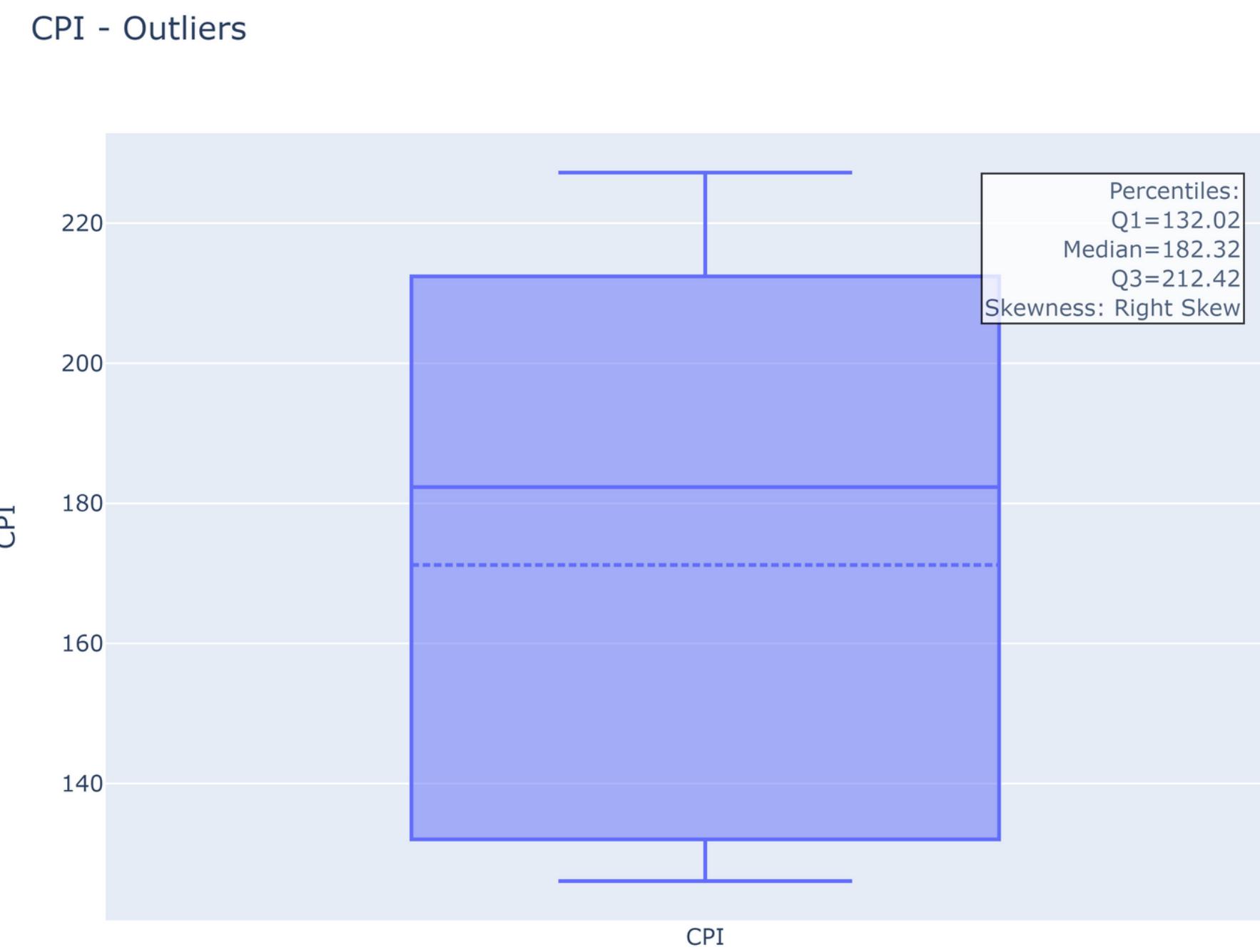
Consumer Price Index (CPI)



Distribution



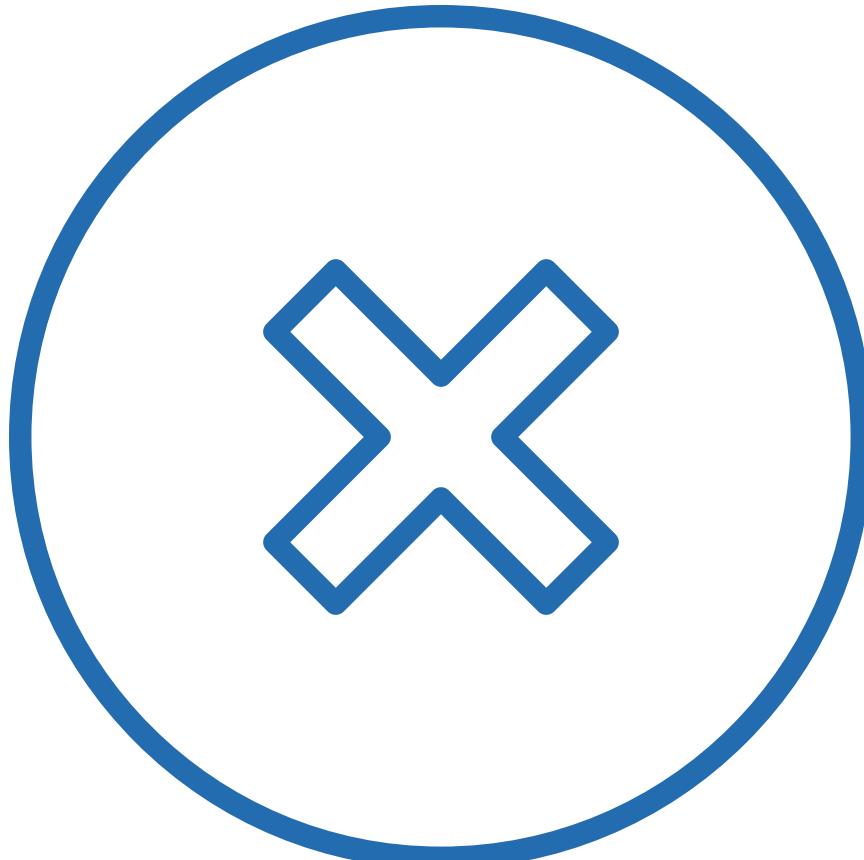
Box Plot - Outliers



Hypothesis 4

Null Hypothesis (H₀): The mean weekly sales during periods of high CPI are not different from the mean weekly sales during periods of low CPI.

Alternative Hypothesis (H_a): The mean weekly sales during periods of high CPI are significantly different from the mean weekly sales during periods of low CPI



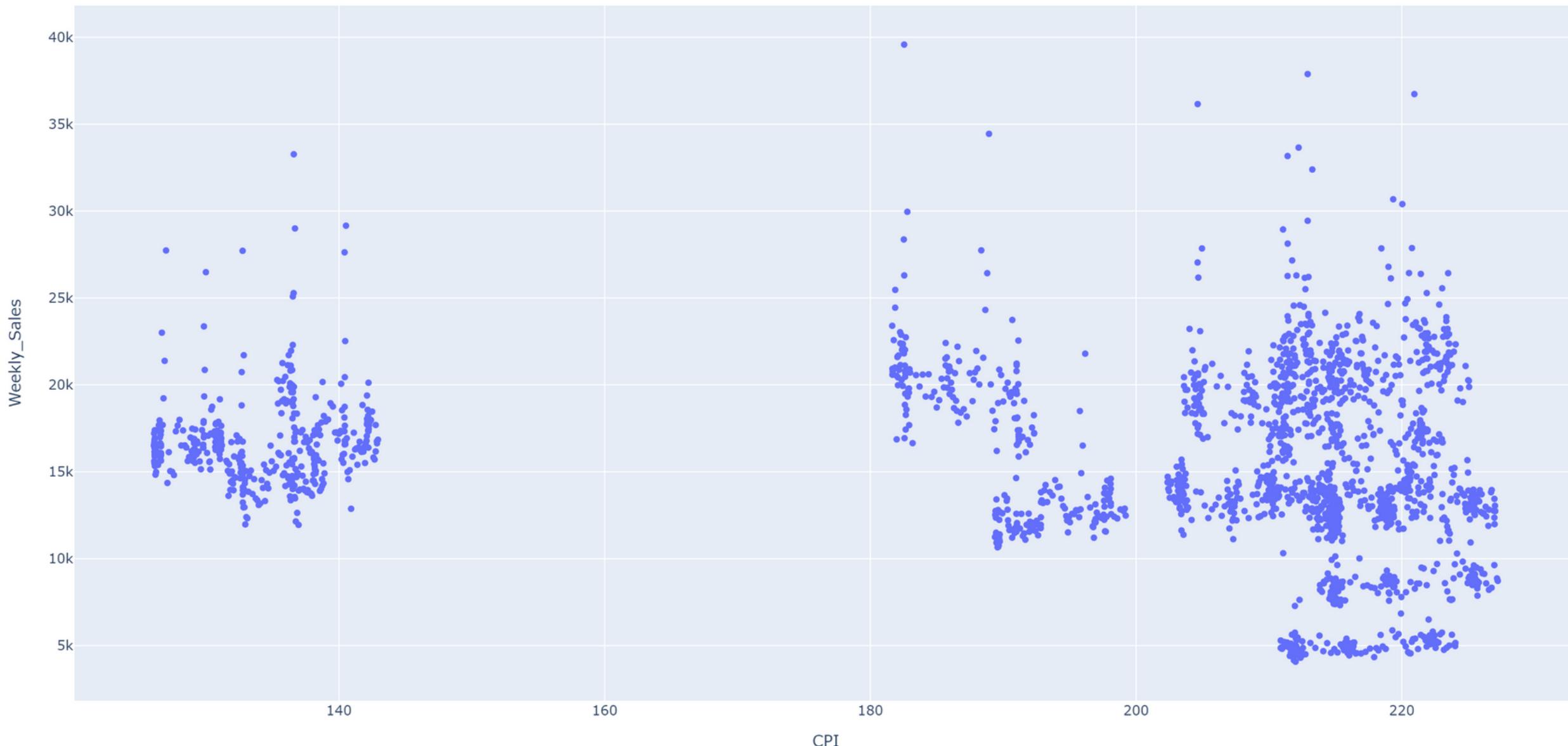
Reject null hypothesis: The mean weekly sales during periods of high CPI are significantly different from the mean weekly sales during periods of low CPI.

Significance level = 0.05

p_value = 1.68766×10^{-30} z_statistic = -11.478

CPI vs Weekly Sales

Average Weekly Sales by Consumer Price Index (CPI)



This figure shows how weekly sales are decreasing obviously when CPI is increasing, So that there is a real difference between weekly sales during high and low CPI periods.

Unemployment Rate

Distribution

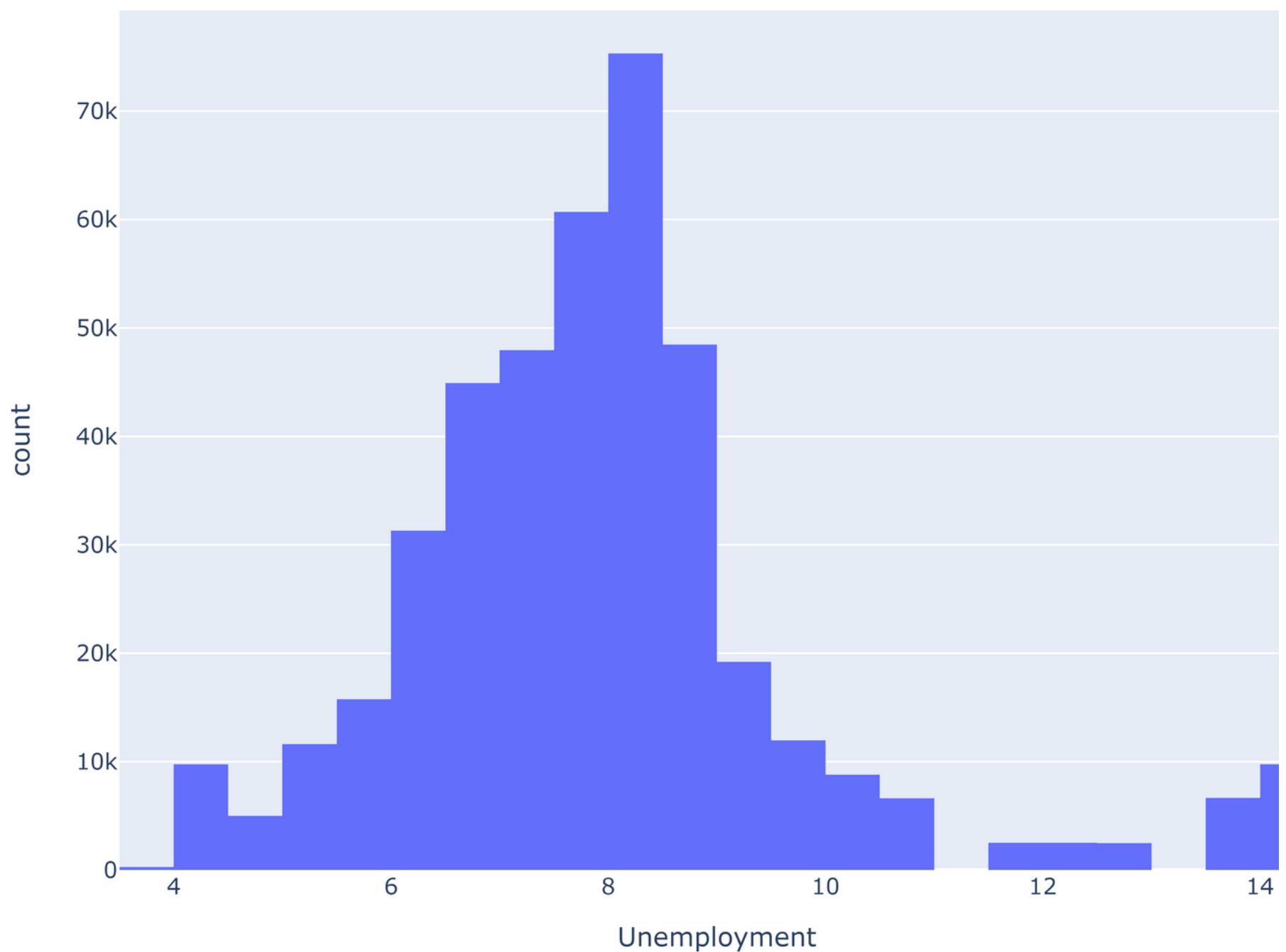
Box Plot -
Outliers

Hypothesis

Average
Weekly
Sales

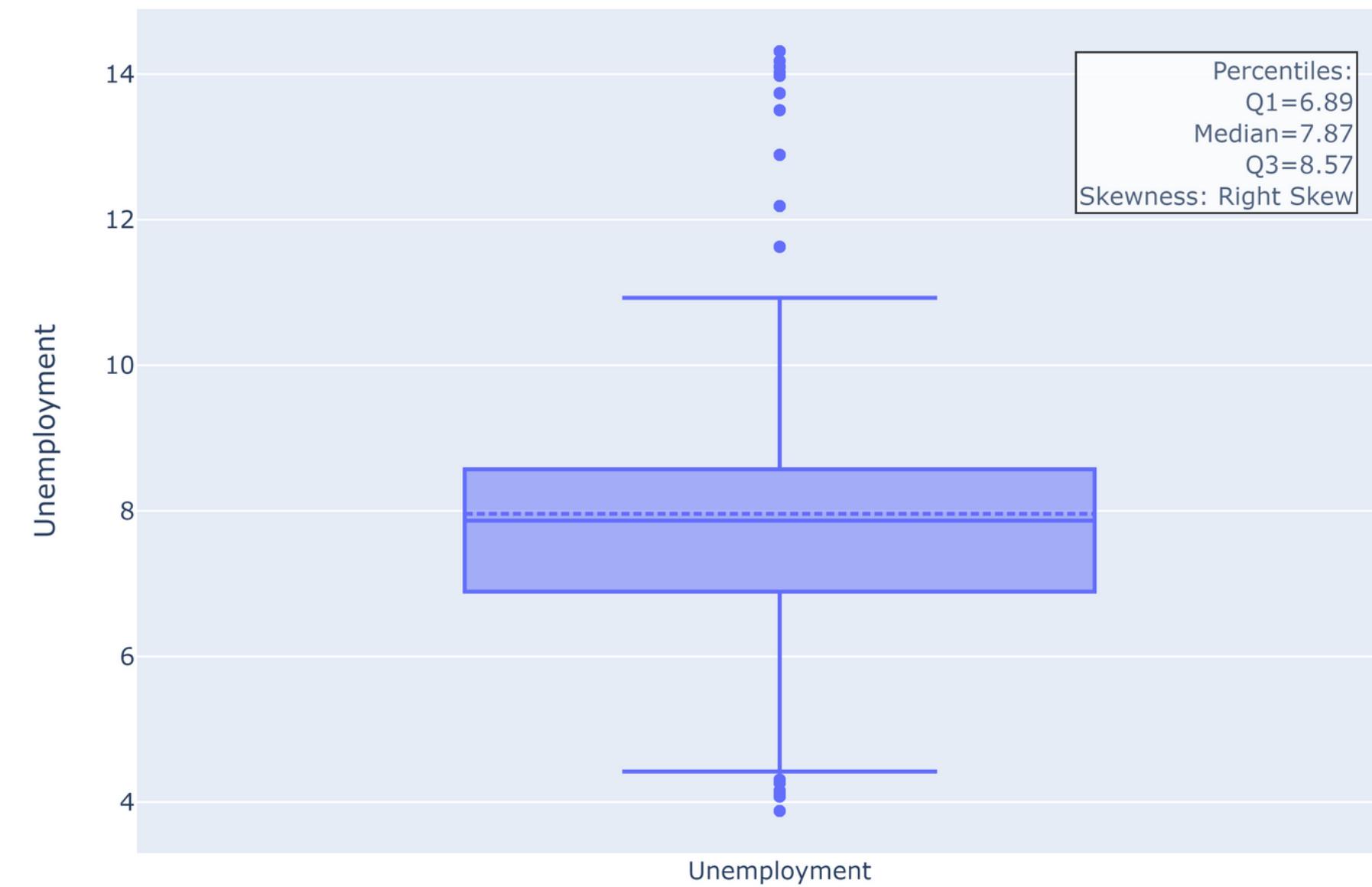
Distribution

Unemployment



Box Plot - Outliers

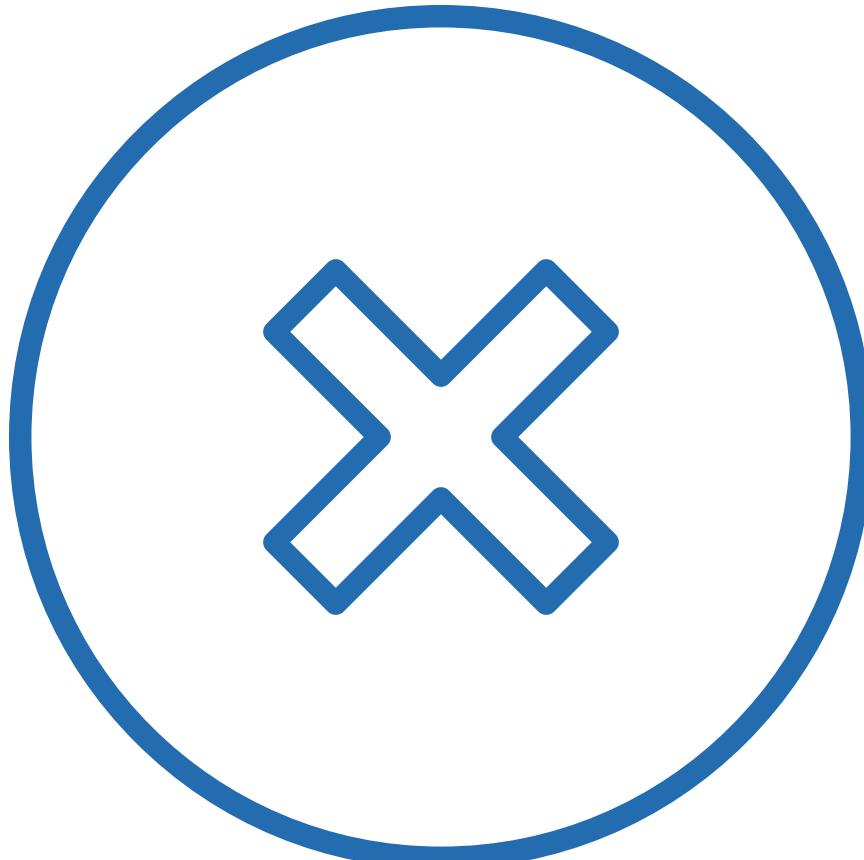
Unemployment - Outliers



Hypothesis 5

Null Hypothesis (H₀): No significant difference in weekly sales between high and low unemployment periods.

Alternative Hypothesis (H_a): There is a significant difference in the mean weekly sales between periods of high unemployment rates and periods of low unemployment rates.



Reject null hypothesis: There is a significant difference in weekly sales between high and low unemployment periods.

Significance level = 0.05

p_value = 1.83774×10^{-20} z_statistic = -9.2713

Unemployment Rate vs Weekly Sales

Average Weekly Sales by Unemployment rate



Average weekly sales are strongly affected by the unemployment rate. It is almost going to be immutable by the increasing of the unemployment rate.

Temperature

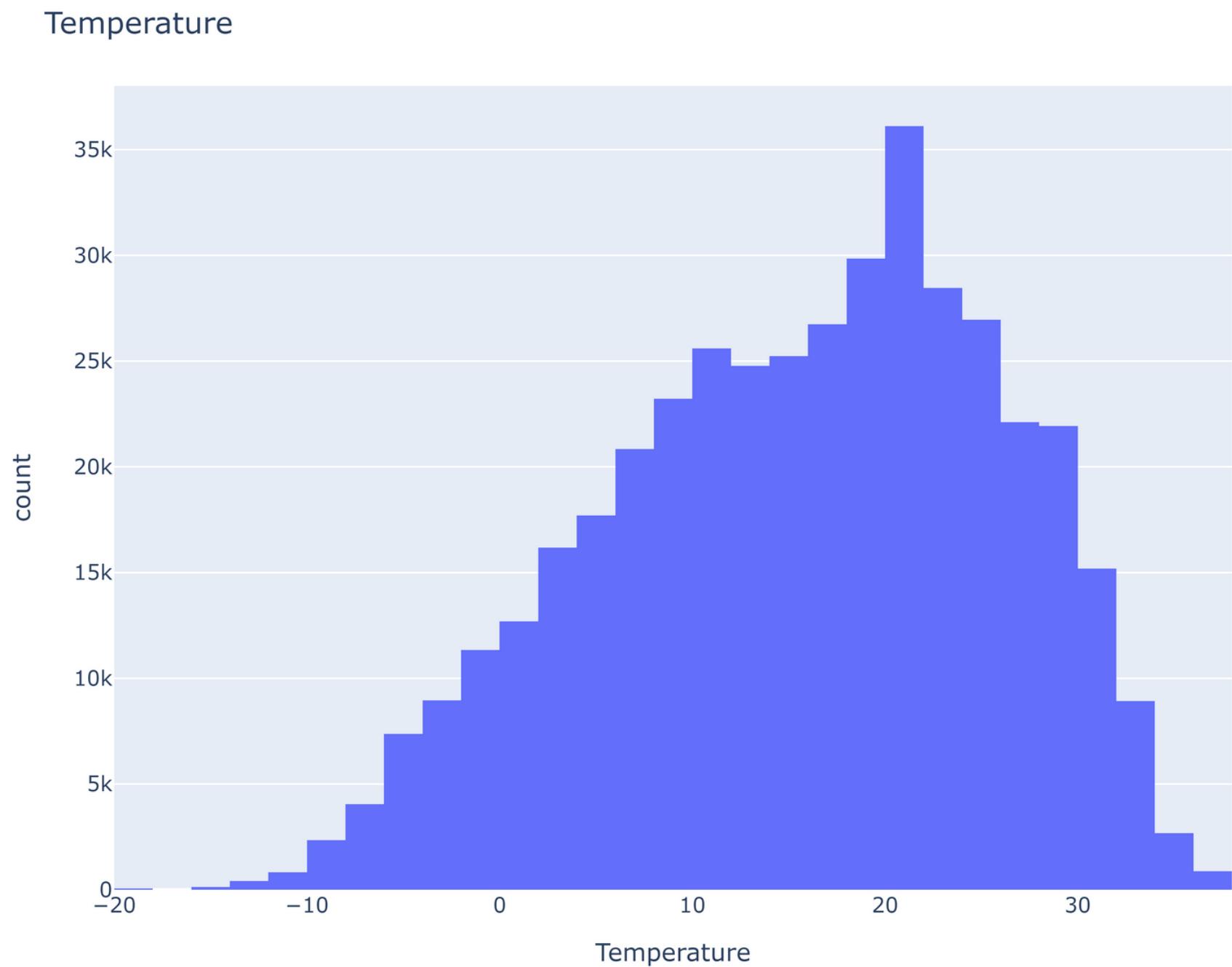
Distribution

Box Plot -
Outliers

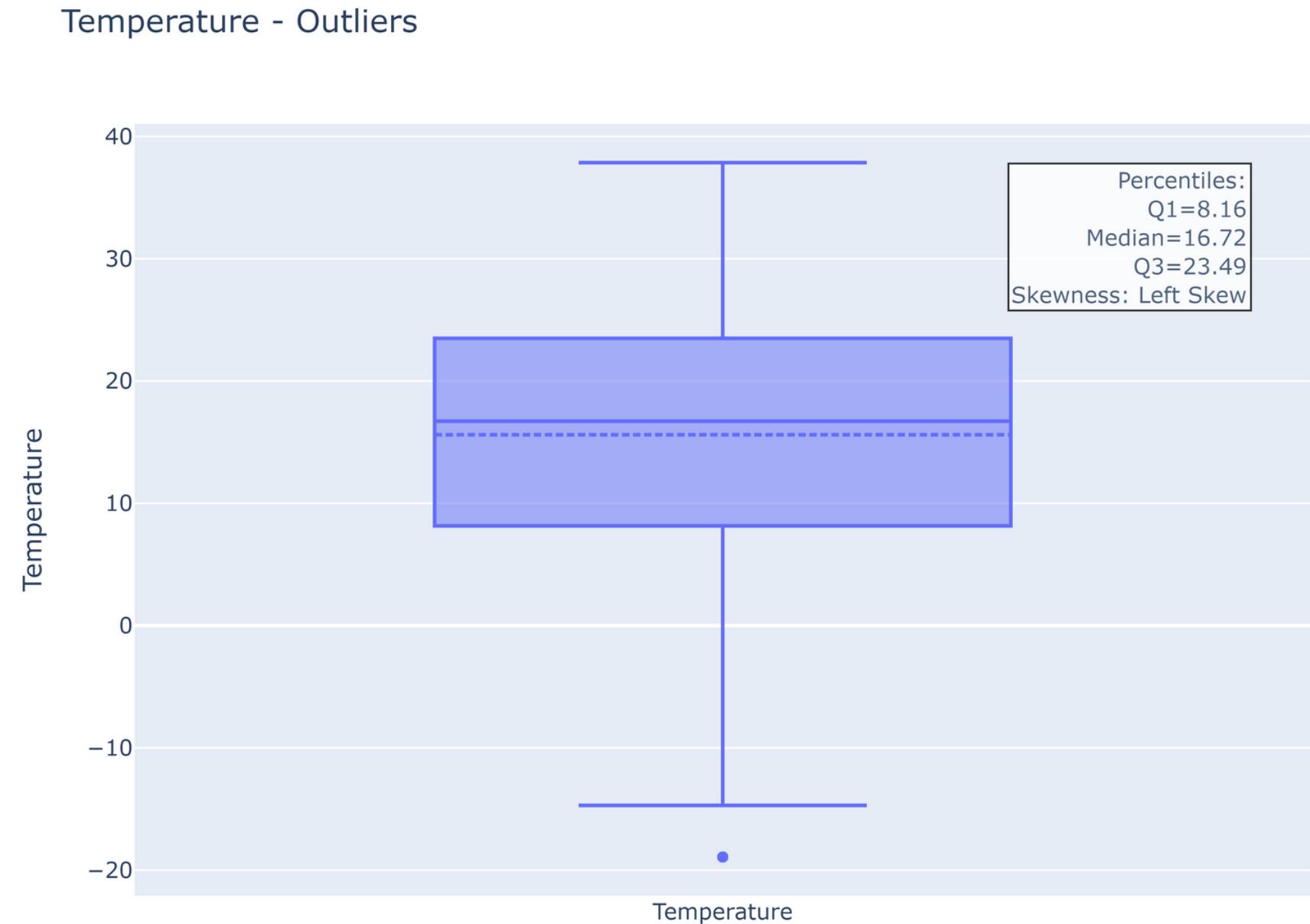
Average
Weekly
Sales

Hypothesis

Distribution



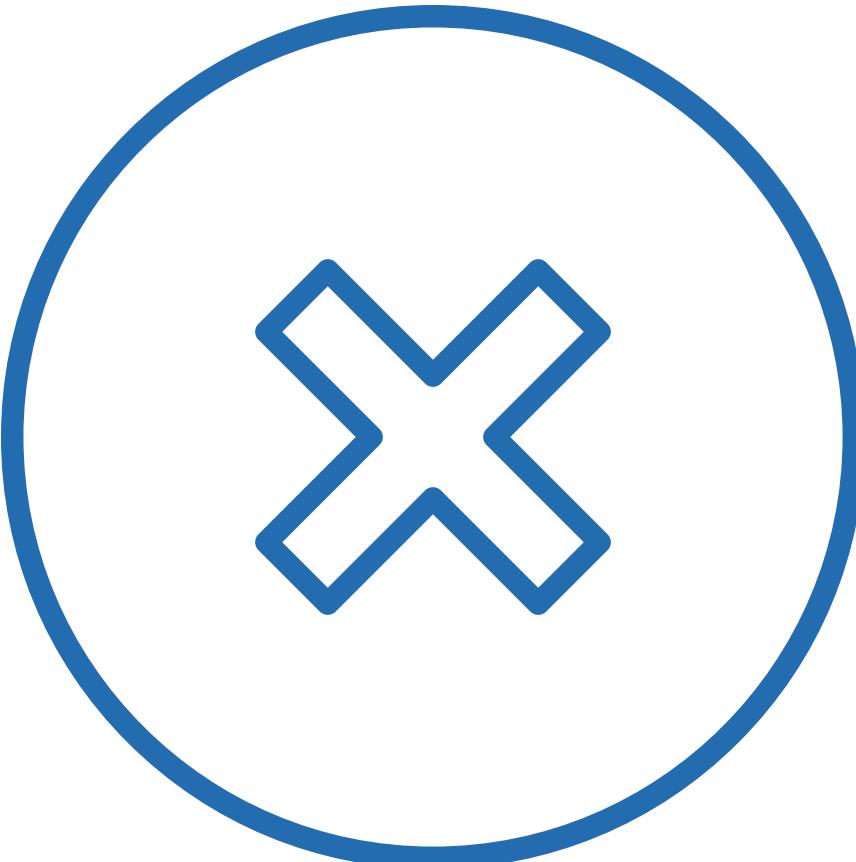
Box Plot - Outliers



Hypothesis 6

Null Hypothesis (H₀): There is no significant variance in weekly sales between high temperature periods and low temperature periods.

Alternative Hypothesis (H_a): The average weekly sales during high temperature periods show a significant difference from the average weekly sales during low temperature periods.



Fail to reject null hypothesis: No significant evidence suggests a difference in average weekly sales between periods of high temperatures and periods of low temperatures.

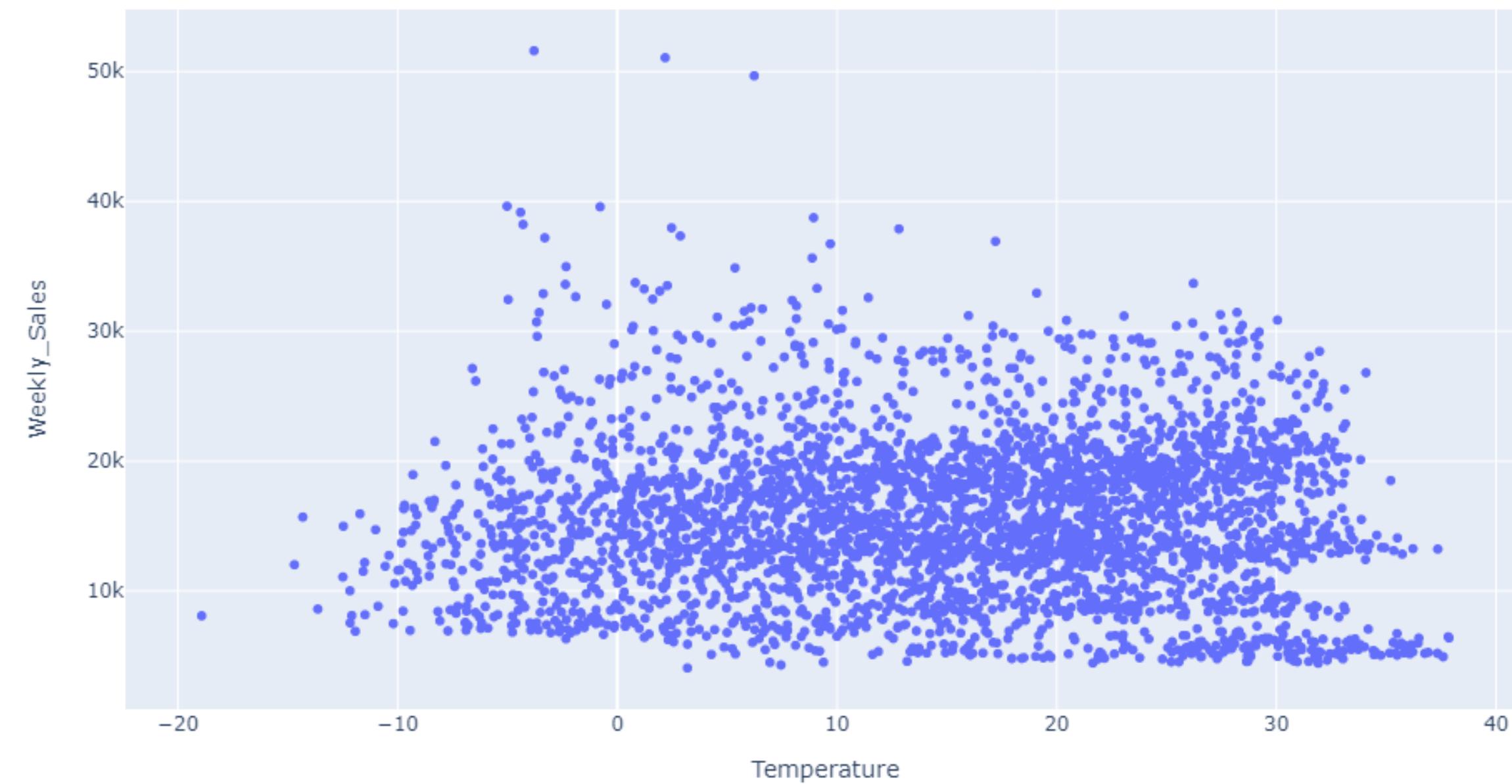
Significance level = 0.05

p_value = 0.16435

z_statistic = -1.3905

Temperature vs Weekly Sales

Average Weekly Sales by Temperature



There is no significant relationship between temperatures and weekly sales, except that the real difference is seen in a few cases in extremely high and extremely low temperatures.

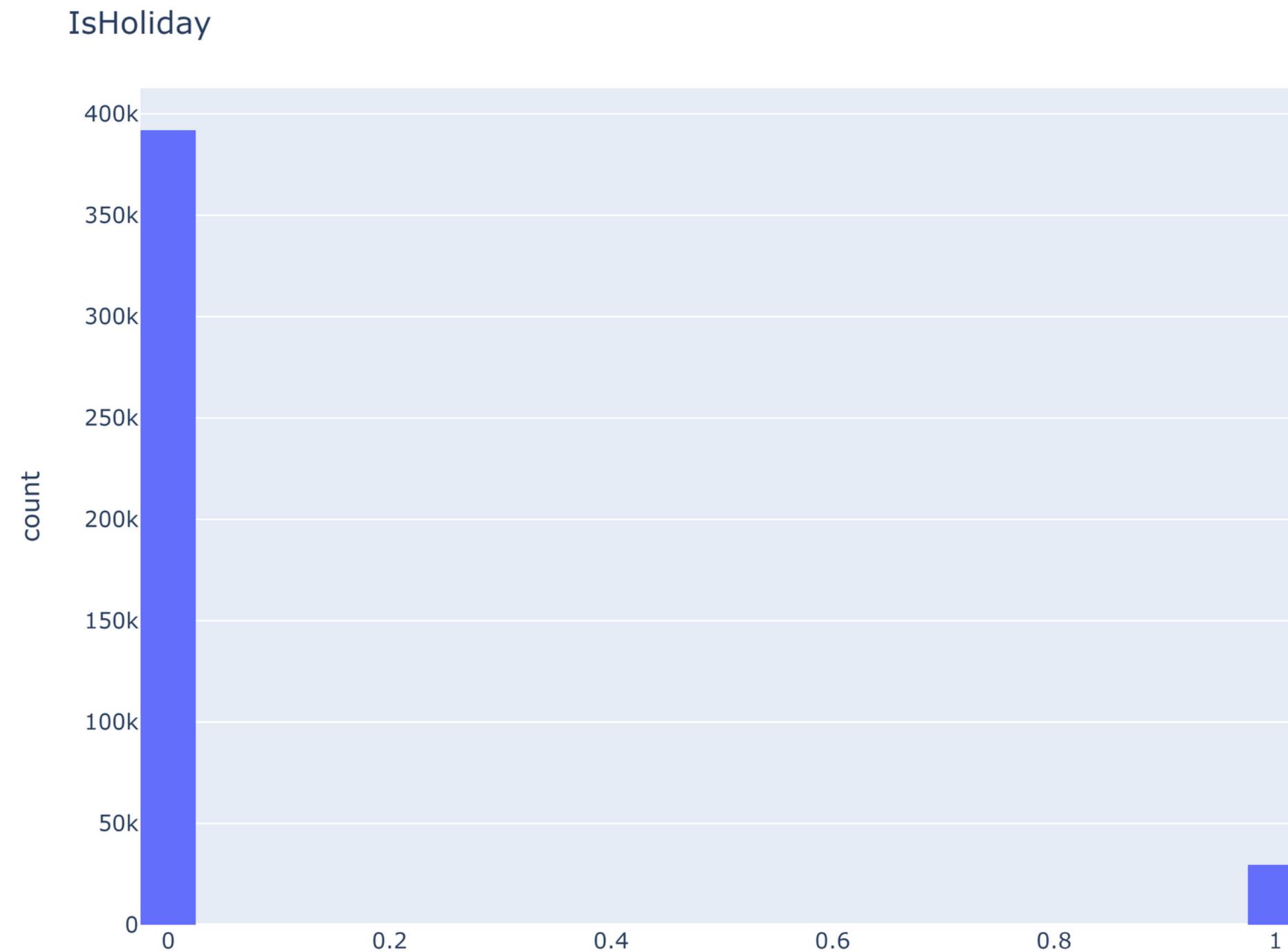
Is it a Holiday

Distribution

Average
Weekly
Sales

Hypothesis

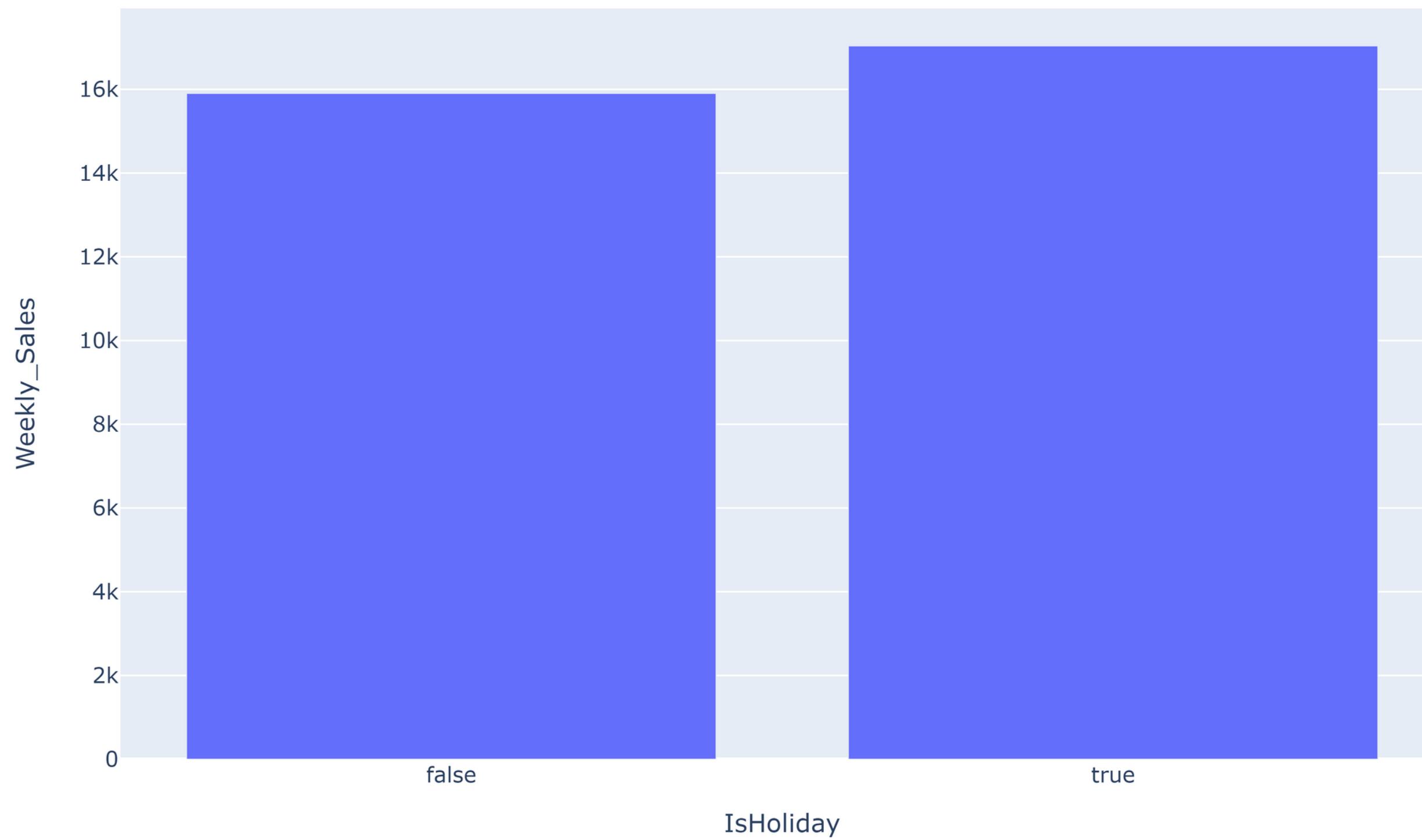
Distribution



The percentage of the holidays are not more than 7.57% from the non holidays.
So does the holidays really affect the weekly sales?
that's what we will see in next slides

Is it a Holiday vs Weekly Sales

Average Weekly Sales by Holiday



Is it a Holiday

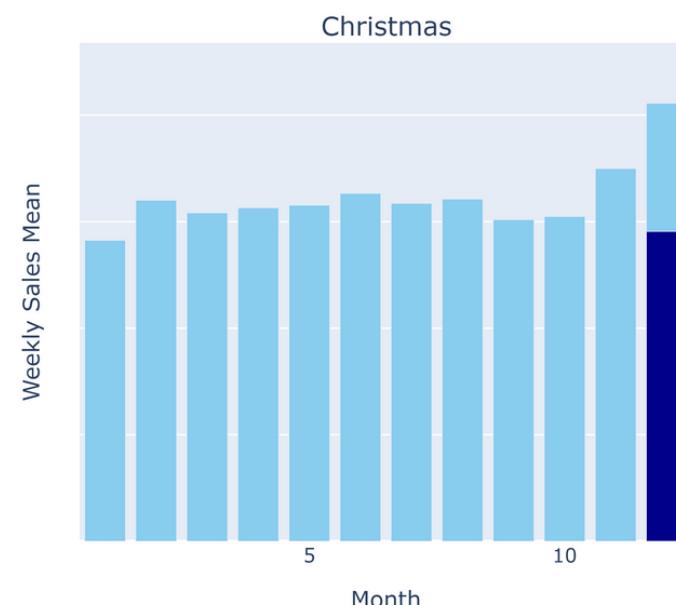
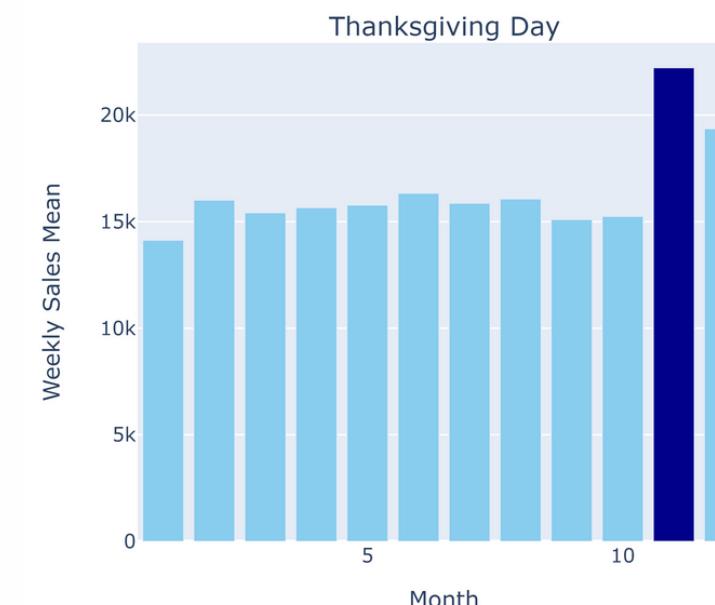
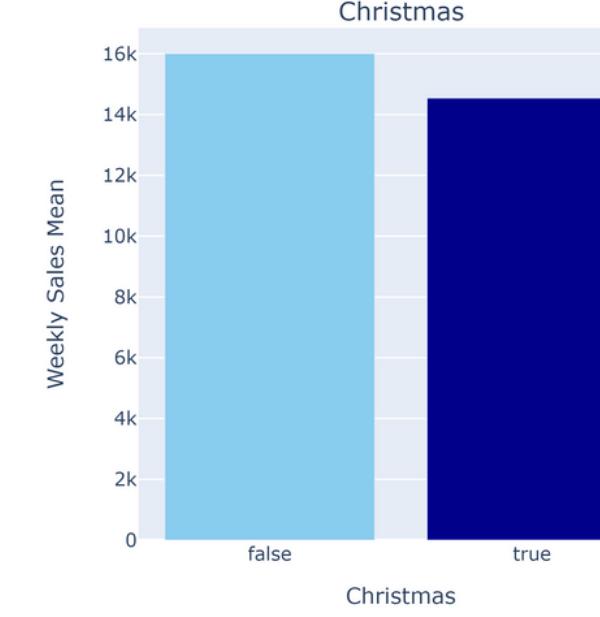
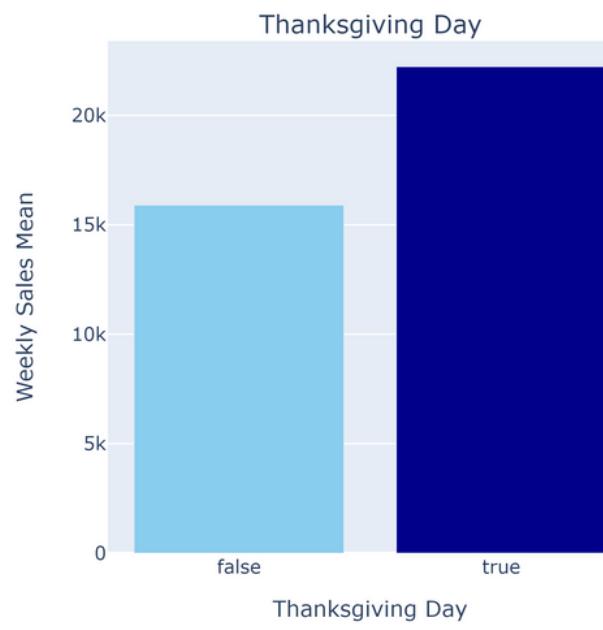
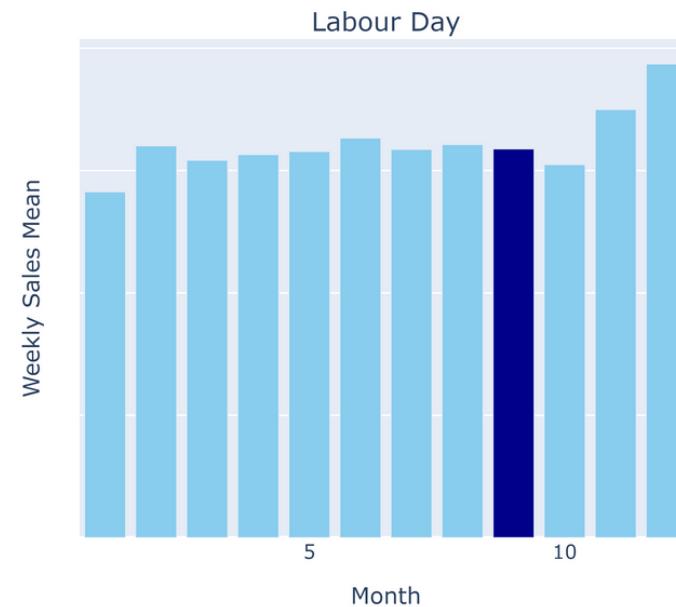
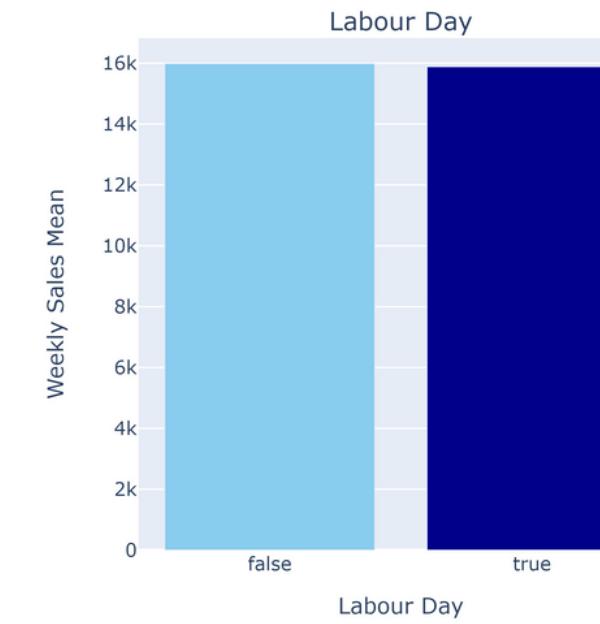
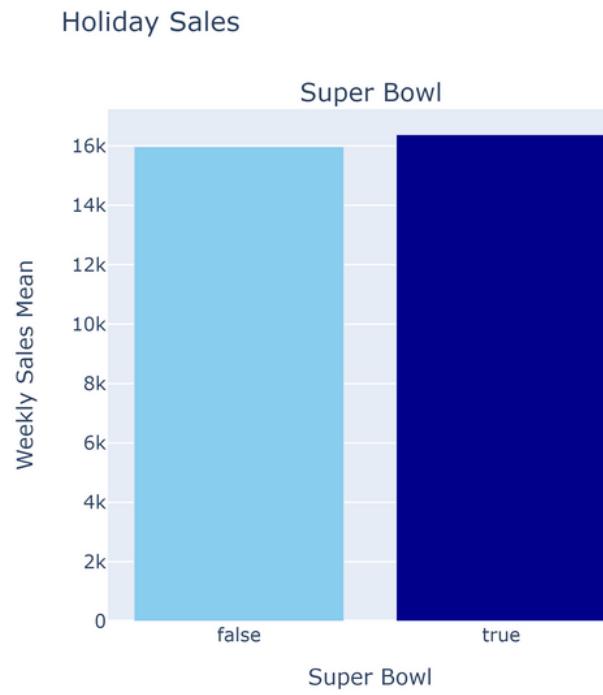
Super Bowl

Labour Day

Thanksgiving's
Day

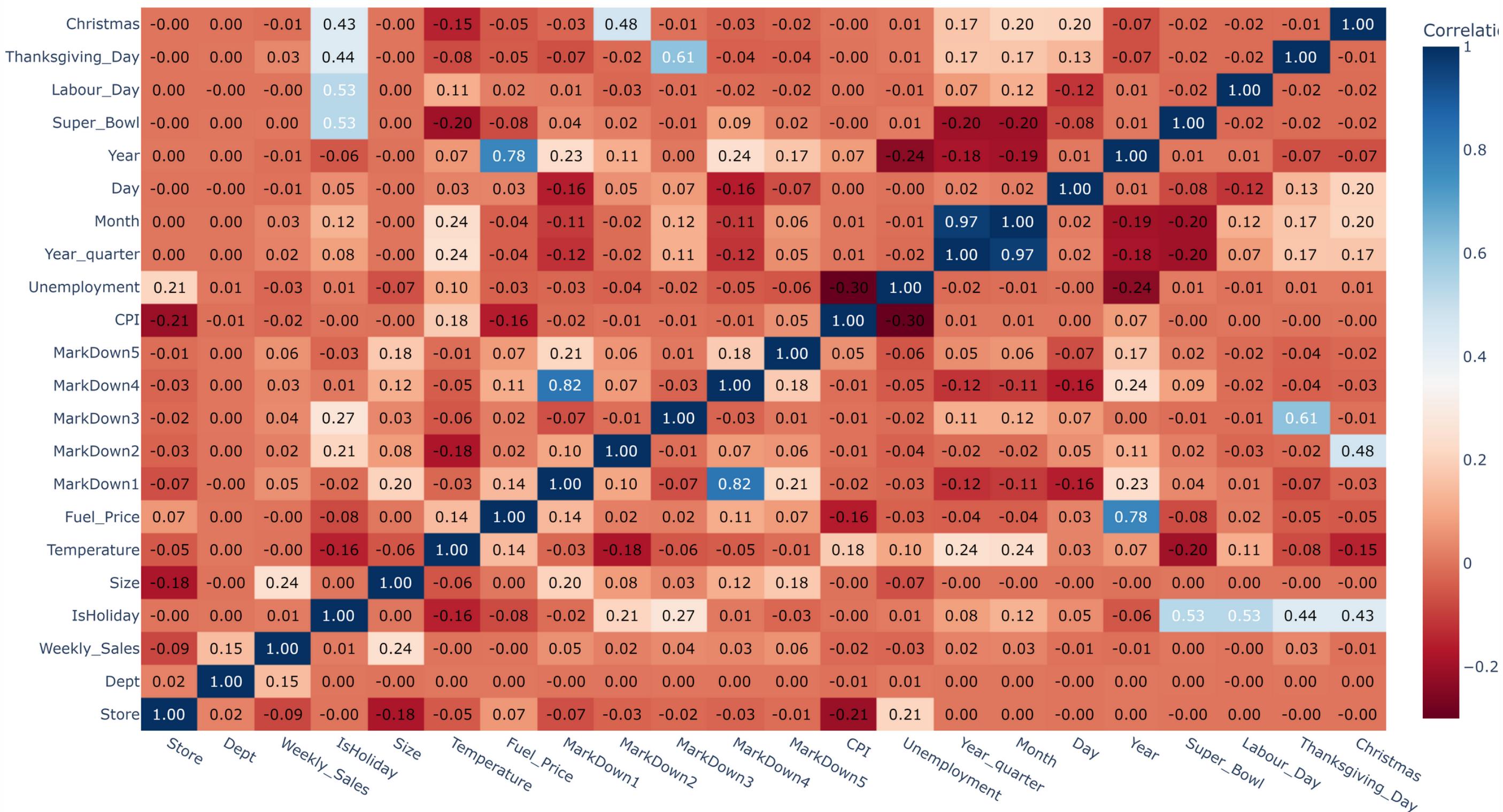
Christmas

Weekly Sales vs Holidays and Month



Correlation

Correlation Heatmap



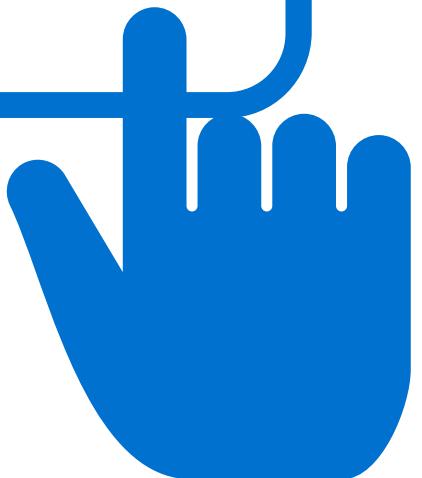
Before Feature Selection

- Data Cleansing
- Outlier treatment
- Feature Engineering
- Model Selection

FEATURE ENGINEERING

RFE

MAE 2982



Information
Criterion

MAE 3493.4

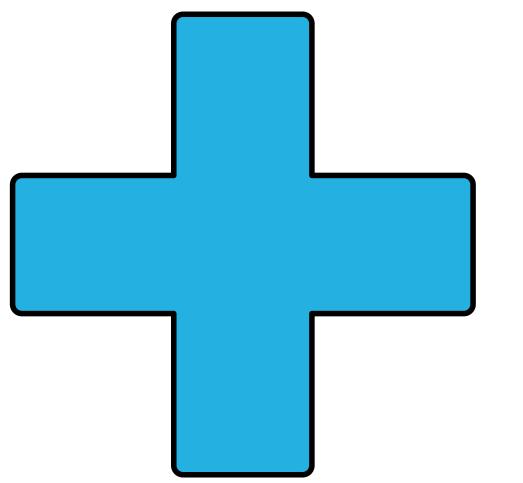
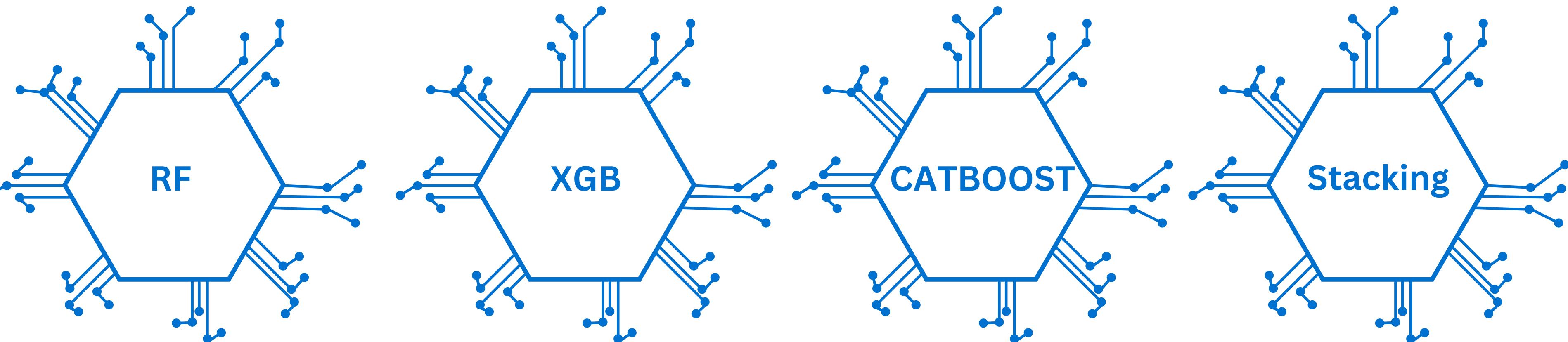
Premutation
Importance

MAE 3351.1

Mutual
Information

MAE 3133.3

MODEL TRAINING



GRIDSEARCHCV

MODEL TRAINING

Model Name	Training Data (r2)	Testing Data (r2)	MAE
RandomForestRegressor	0.9722	0.9612	2264.9
XGBRegressor	0.9480	0.9455	3063.3
CatBoostRegressor	0.9696	0.9602	2591.3
StackingRegression	0.9736	0.9639	2143.7