

Accelerate Business Growth of Industries with Best-In-Class Self-Service Vertical Use case Builder.

Business Solution for Sales Forecasting



[Business Objective of the Case Study](#)

[Challenges](#)

[Solution Approach](#)

[Implementation Process](#)

[Dataset Definition](#)

[Data Profiling](#)

[Exploratory Data Analysis](#)

[EDA Workflow View](#)

[Key EDA Insights](#)

[Feature Engineering](#)

[Machine Learning Workflow](#)

[Classification ML Model](#)

[Model Registry](#)

[Model Inference](#)

[AutoML](#)

[Business Apps](#)

[Business Impact](#)

[References](#)

Business Objective of the Case Study



- Forecasting the quantity of items for each Product Category in a Market Segment helps in smooth management of the inventory, optimizing shelf life and maximizing the sales of the items.

Challenges



- In this unprecedented time of uncertainty, businesses need to rely on forecasts, now more than ever, to drive their entire growth strategy.
- Smooth Resource allocation and planned management of cash flow is a major challenge
- Its extremely important to manage the budget under inflation and downward economic situation
- Investors need to visualize the sales performance and goals.

Solution Approach

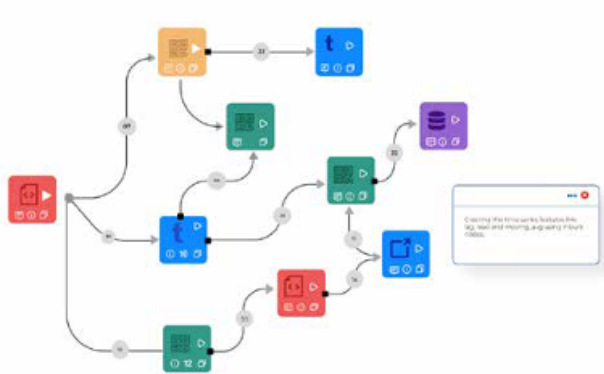


- Quantitative Sales Forecasting involves statistical analysis of past sales data to forecast. It is based on objective calculations and yields more accurate results.
- Past sales performance is a good leading indicator of future sales performance. Historical conversion rates tell you how many prospects, teams, or individuals were able to convert over a given period of time.
- Automatically capturing sales activity and customer engagement data.
- Using Machine learning and AI to analyze that deal data to provide you with meaningful and actionable insights for better forecast accuracy and efficiency. ARIMA is one of the best-performing algorithms used for time series forecasting. In the workflow below, we use ARIMA to forecast the number of items that would be sold for a Category in a Market Segment.
- Real-time forecasting that gives you better accuracy with up-to-the minute opportunity data and saves you time by avoiding telephone calls.

Implementation Process

- Dataset Definition
- Exploratory Data Analysis
- Data Enrichment
- Feature Engineering

- Customer Clustering
- Cluster Evaluation



Dataset Definition

We have used the dataset available at Kaggle which was made available by Walmart. This can be downloaded from [Walmart Store Sales Forecasting | Kaggle](#)

Stores: It provides the type and size

Sales: It contains information like for a store, its weekly sales, whether the week is a special holiday week.

Features: This file holds all the main features which need to be analyzed for sales. This contains data for a store on a particular day, the area temperature, whether that day was a holiday.

Exploratory Data Analysis



Powerful Exploratory Data Analysis Workflow can be designed using the Fire Insight Studio. The workflow provides crucial insights in graphical form.

Key EDA Insights

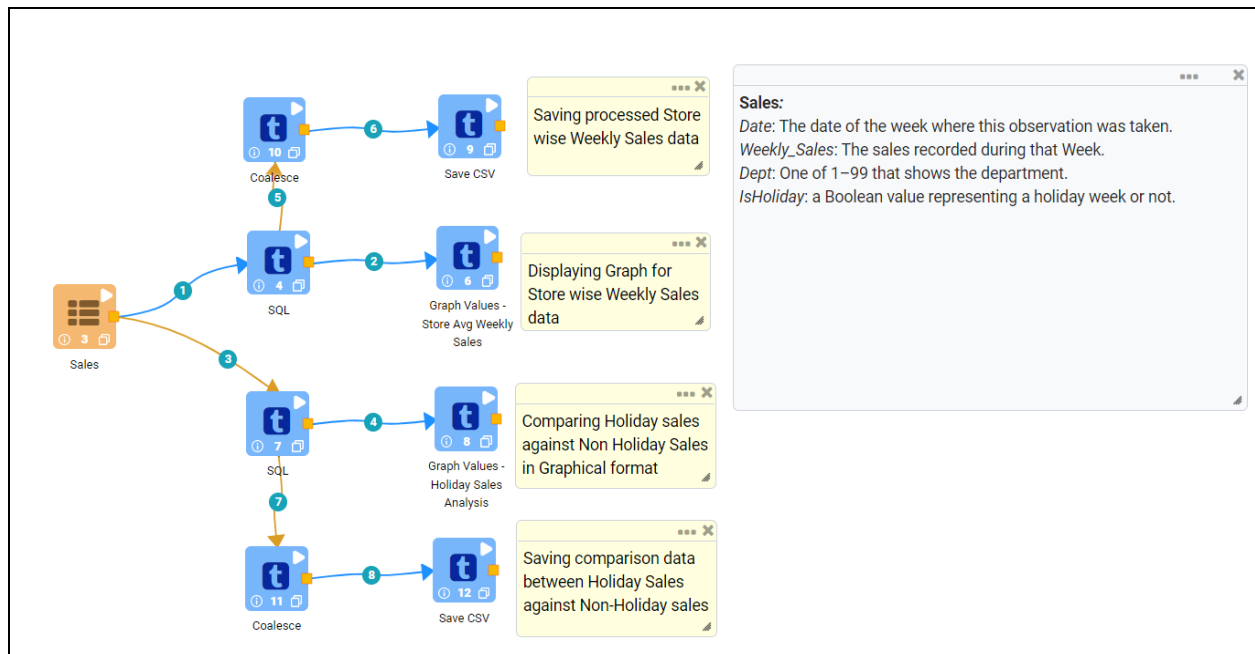


Weekly Sales by Store

Workflow performs following tasks:

1. Reads Sales dataset
2. Processes Store wise Weekly Sales data
3. Processes comparison between Holiday and Non-Holiday sales
4. Saves data in CSV format
5. Displays processed data in Graphical format

EDA Workflow Design

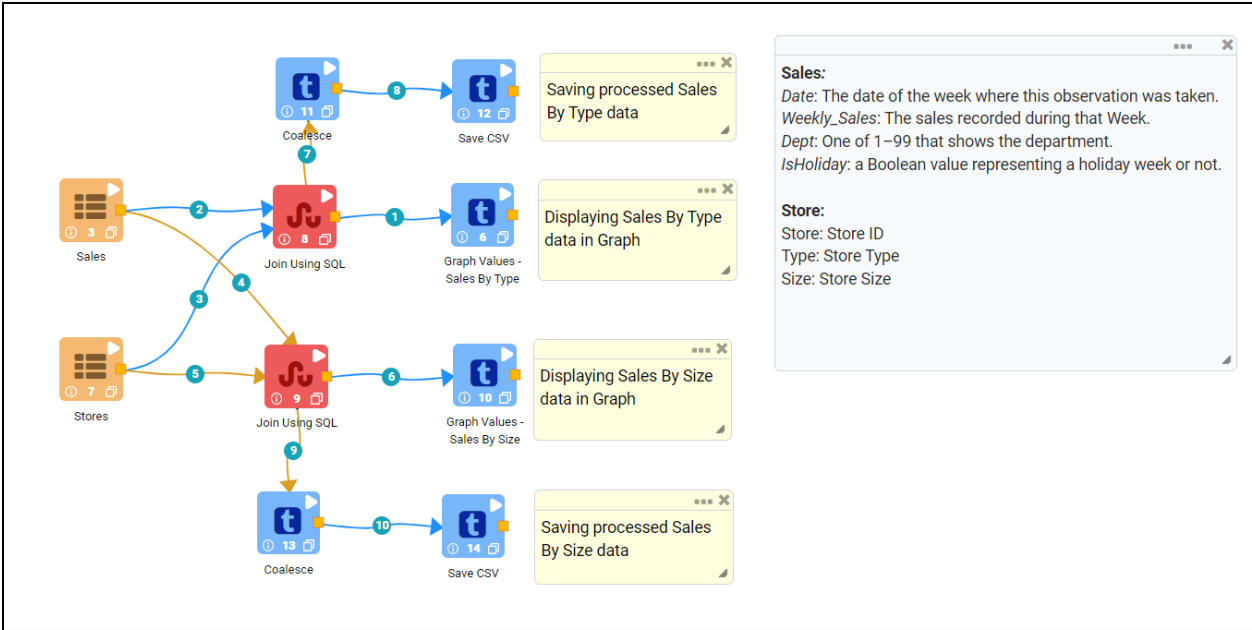


Sales By Type and Size

Workflow performs following tasks:

1. Reads Sales and Stores dataset
2. Processes Sales By Type data
3. Processes Sales By Size
4. Saves data in CSV format
5. Displays processed data in Graphical format

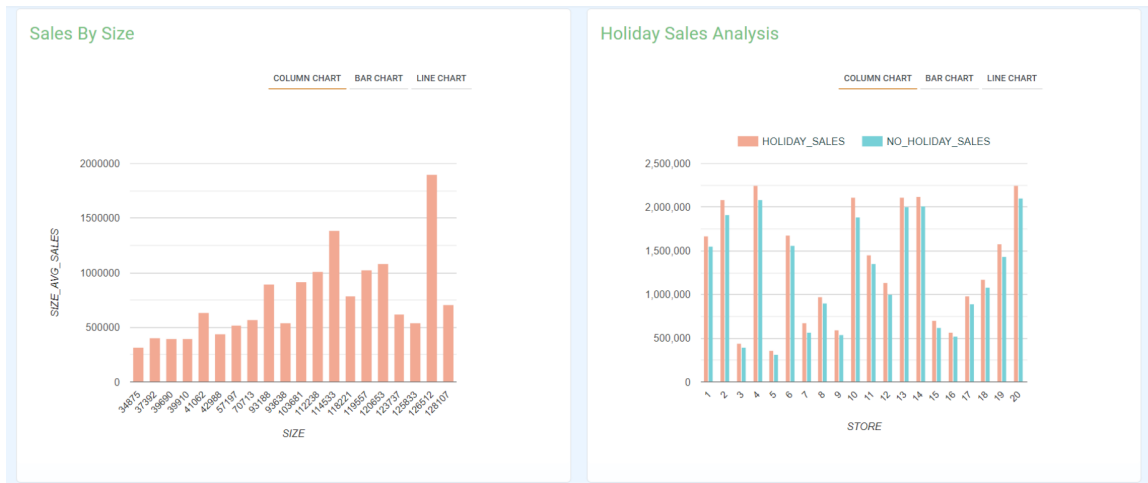
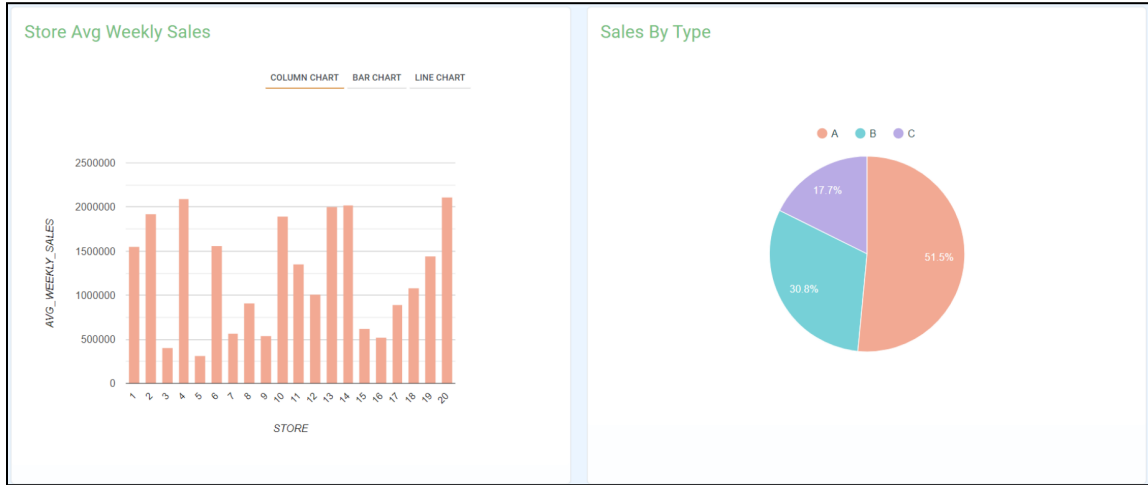
EDA Workflow Design



Report displays following KPIs:

1. Store Weekly Sales
2. Store Sales By Type Analysis
3. Store Sales By Size Analysis
4. Store Sales during Holiday Vs Non-Holiday Analysis

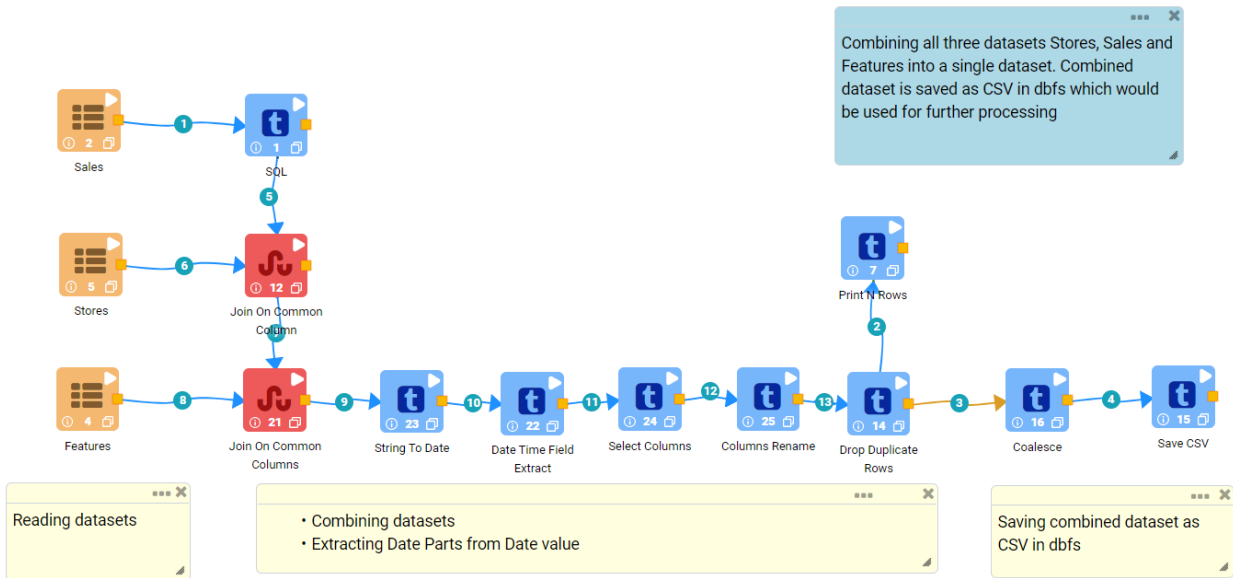
EDA Insights



Data Preparation

Data Pre-processing: Combining Datasets

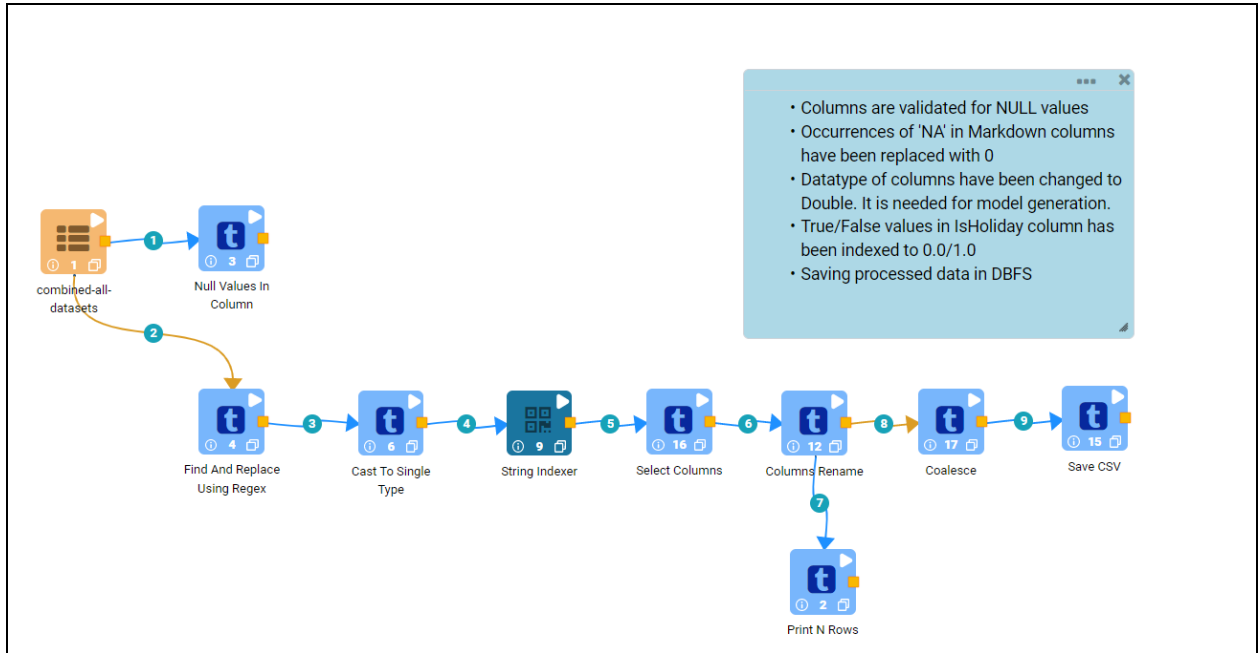
Workflow View



Sample Dataset

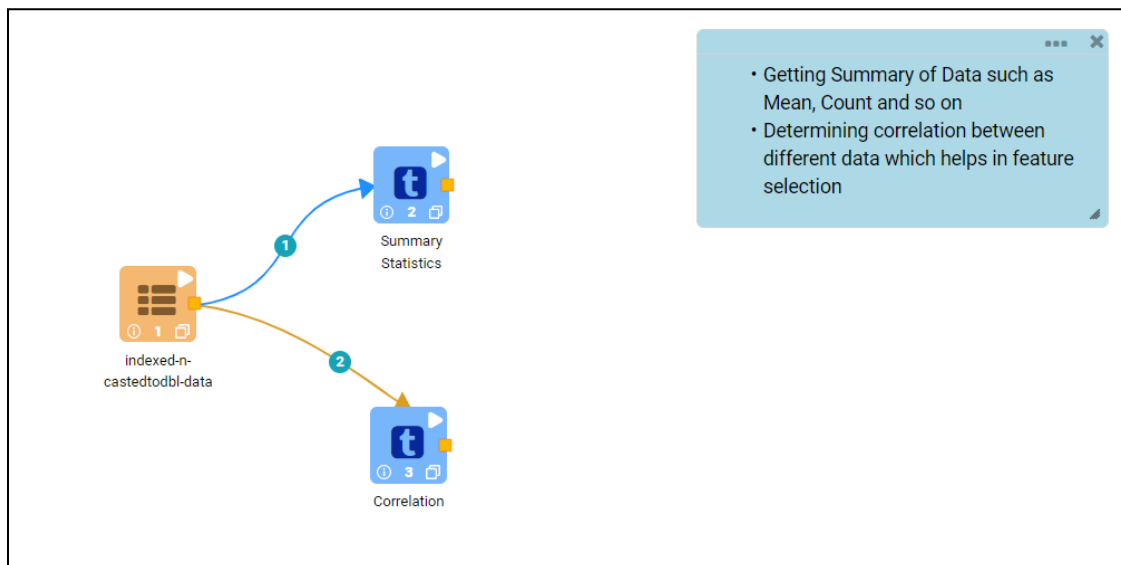
Store	Date	IsHoliday	Weekly_Sales	Type	Size	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	Year/Value	Month/Value	DayOf/Value	Wei
1	2010-03-05	FALSE	1554806.68	A	151315	46.5	2.625	NA	NA	NA	NA	NA	211.3501429	8.106	2010	3	5	9
1	2010-04-02	FALSE	1594968.2799999998	A	151315	62.27	2.719	NA	NA	NA	NA	NA	210.8204499	7.808	2010	4	2	13
1	2011-12-16	FALSE	1881176.67	A	151315	51.63	3.159	5011.32	67	347.37	225.79	4011.37	219.1794533	7.866	2011	12	16	50
2	2010-03-26	FALSE	1750197.81	A	202307	51.26	2.732	NA	NA	NA	NA	NA	210.6766095	8.324	2010	3	26	12
2	2012-03-30	FALSE	1857480.8399999996	A	202307	66.83	3.845	14369.3	164.77	17	1017.31	3470.49	220.9991248	7.057	2012	3	30	13
2	2012-07-27	FALSE	1757923.88	A	202307	86.37	3.407	6999.14	197.72	2.53	6319.33	5007.01	221.5785461	6.565	2012	7	27	30
3	2012-10-26	FALSE	405432.70000000007	B	37392	74.66	3.506	370.99	6.35	4	5.94	180.07	226.9873637	6.034	2012	10	26	43
4	2010-04-30	FALSE	1817273.28	A	205863	53.04	2.787	NA	NA	NA	NA	NA	126.3805667	7.896	2010	4	30	17
4	2011-12-16	FALSE	2207742.1300000004	A	205863	64.79	3.361	NA	NA	NA	NA	NA	129.7821613	5.143	2011	10	21	42

Data Validation and Indexing



Feature Engineering

Feature Selection and Correlation

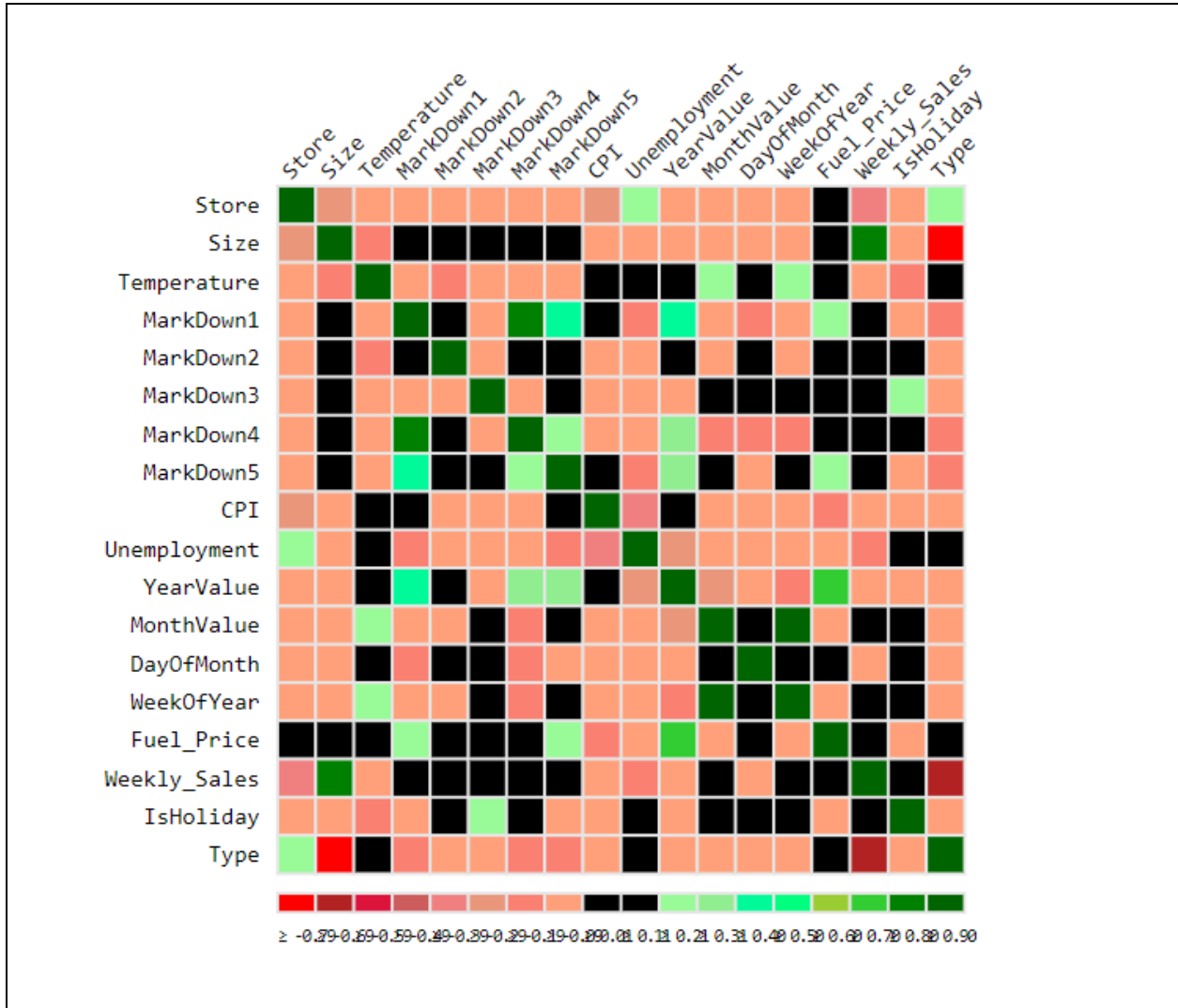


Summary

summary	store	size	temperature	markdown1	markdown2	markdown3	markdown4	markdown5	cpi	unemployment	yearvalue
count	6435	6435	6435	6435	6435	6435	6435	6435	6435	6435	6435
mean	23.0	130287.6	60.664	2429.019	818.873	429.184	1008.87	1581.807	171.578	7.999	2010.965
min	1.0	34875.0	-2.06	0.0	-265.76	-29.1	0.0	0.0	126.064	3.879	2010.0
25_percentile	12.0	70713.0	47.43	0.0	0.0	0.0	0.0	0.0	131.735	6.891	2010.0
50_percentile	23.0	126512.0	62.67	0.0	0.0	0.0	0.0	0.0	182.6165205	7.874	2011.0
75_percentile	34.0	202307.0	74.93	2297.32	0.09	3.69	313.76	1983.03	212.7396889	8.622	2012.0
max	45.0	219622.0	100.14	88646.76	104519.54	141630.61	67474.85	108519.28	227.2328068	14.313	2012.0
stdev	12.988	63117.022	18.445	5871.478	4879.257	5273.523	3753.943	4098.199	39.357	1.876	0.797
variance	168.693	3.983758524889E9	340.216	3.4474250683E7	2.3807153646E7	2.781004665E7	1.4092090401E7	1.6795231274E7	1548.951	3.519	0.635

From The Correlation Matrix, we can see that :

1. There exists a strong relationship between the size of the store and the holiday period.
2. There exists a relationship between temperature and the Month which can be analyzed further to have store specific sales/markdowns.

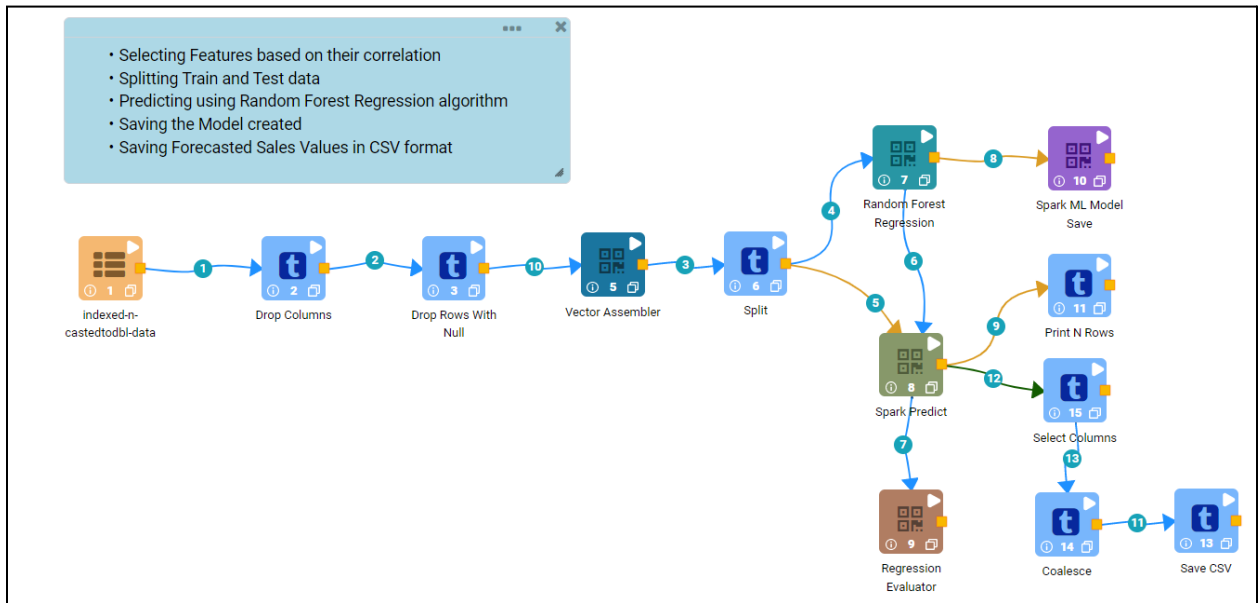


Machine Learning Workflow

Model Building and Forecasting

Random Forest Classifier to predict Churn

- This



Feature Importance

Features Importance

features	Importance %
Size	58.71949499134866
MarkDown1	1.344905397362714
MarkDown2	0.14627810057711887
MarkDown3	3.1231031238630003
MarkDown4	0.9829140935453088
MarkDown5	3.0370018103731735
CPI	4.444333752211314
Unemployment	3.2588895527757225
MonthValue	0.2386855401075589
DayOfMonth	0.26657361710715055
WeekOfYear	1.7414382567311766

Model Inference

Regression Evaluator

Metrics

rmse: 257197.88898000846

mse: 6.615075409577275E10

mae: 188805.17243531195

Model Inference

Predicted Result

Month	WeekOfYear	Fuel_Price	Weekly_Sales	IsHoliday	Type	Vector_Data	Predicted_Weekly_Sales
Type	DoubleType	DoubleType	DoubleType	DoubleType	DoubleType	org.apache.spark.ml.linalg.VectorUDT@3bfc3ba7	DoubleType
13.0	2.719	1594968.28	0.0	0.0	[16,0.6,7.8,9,10,11,14,15],[151315.0,210.8204499,7.808,4.0,2.0,13.0,2.719,62.27,2010.0]	1021024.279	
50.0	3.159	1881176.67	0.0	0.0	[151315.0,5011.32,67.0,347.37,225.79,4011.37,219.1794533,7.866,12.0,16.0,50.0,3.159,0.0,0.51,63,2011.0]	1356223.551	
43.0	3.506	405432.7	0.0	1.0	[37392.0,370.99,6.35,4.0,5.94,180.07,226.9873637,6.034,10.0,26.0,43.0,3.506,0.0,1.0,74.66,2012.0]	458908.637	
43.0	3.506	405432.7	0.0	1.0	[37392.0,370.99,6.35,4.0,5.94,180.07,226.9873637,6.034,10.0,26.0,43.0,3.506,0.0,1.0,74.66,2012.0]	458908.637	
40.0	3.285	1481739.2	0.0	0.0	[16,0.6,7.8,9,10,11,14,15],[202505.0,218.59704,6.551,10.0,7.0,40.0,3.285,74.1,2011.0]	1569519.714	
34.0	3.74	1166479.51	0.0	0.0	[16,0.6,7.8,9,10,11,14,15],[206302.0,129.2832581,13.503,8.0,26.0,34.0,3.74,94.61,2011.0]	1438061.492	
28.0	3.532	1144254.26	0.0	0.0	[16,0.6,7.8,9,10,11,14,15],[203007.0,193.3125484,8.622,7.0,15.0,28.0,3.532,70.55,2011.0]	1553271.327	
4.0	3.038	541037.98	0.0	1.0	[70713.0,747.21,559.46,0.0,47.0,2332.92,196.8262786,8.256,1.0,27.0,4.0,3.038,0.0,1.0,26.9,2012.0]	549415.407	
41.0	3.08	1703850.25	0.0	1.0	[126512.0,0.0,0.0,0.0,0.0,0.0,126.3266774,9.003,10.0,15.0,41.0,3.08,0.0,1.0,75.48,2010.0]	1387129.33	
41.0	3.08	1703850.25	0.0	1.0	[126512.0,0.0,0.0,0.0,0.0,0.0,126.3266774,9.003,10.0,15.0,41.0,3.08,0.0,1.0,75.48,2010.0]	1387129.33	



Business Impact

- Accurate sales forecast allows companies to efficiently allocate resources for future growth and manage their cash flow.
- Sales forecasts help set benchmarks for future trends and allow leaders to course correct early. Revenue leaders can align sales quotas and revenue expectations and optimize for more wins.
- Conveys confidence to the board, and the management team that your business is supported by a reliable forecasting machine that will scale well in the future.
- Sales projections facilitate strategic planning and tell you how soon you will be ready for executing and implementing your plans.

References



<https://www.sparkflows.io/>

<https://docs.sparkflows.io>

<https://www.sparkflows.io/videos>

<https://www.sparkflows.io/data-sheets>