Time Series Market Prediction Project

by Stan Chen on Dec 12, 2019

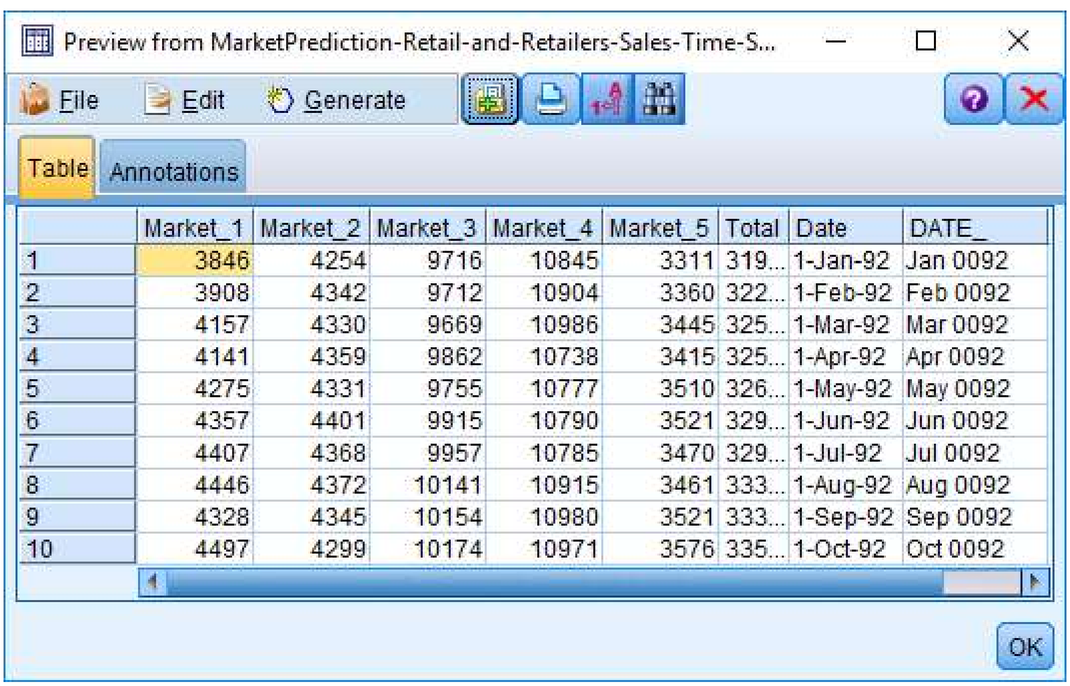
This project is a simple standard time series analysis with IBM SPSS. While other tools and technologies can achieve the exact same results(i.e. KNIME, SageMaker, etc.) , I find the graphs and reports from IBM products concise and easy to explain.

Data source:

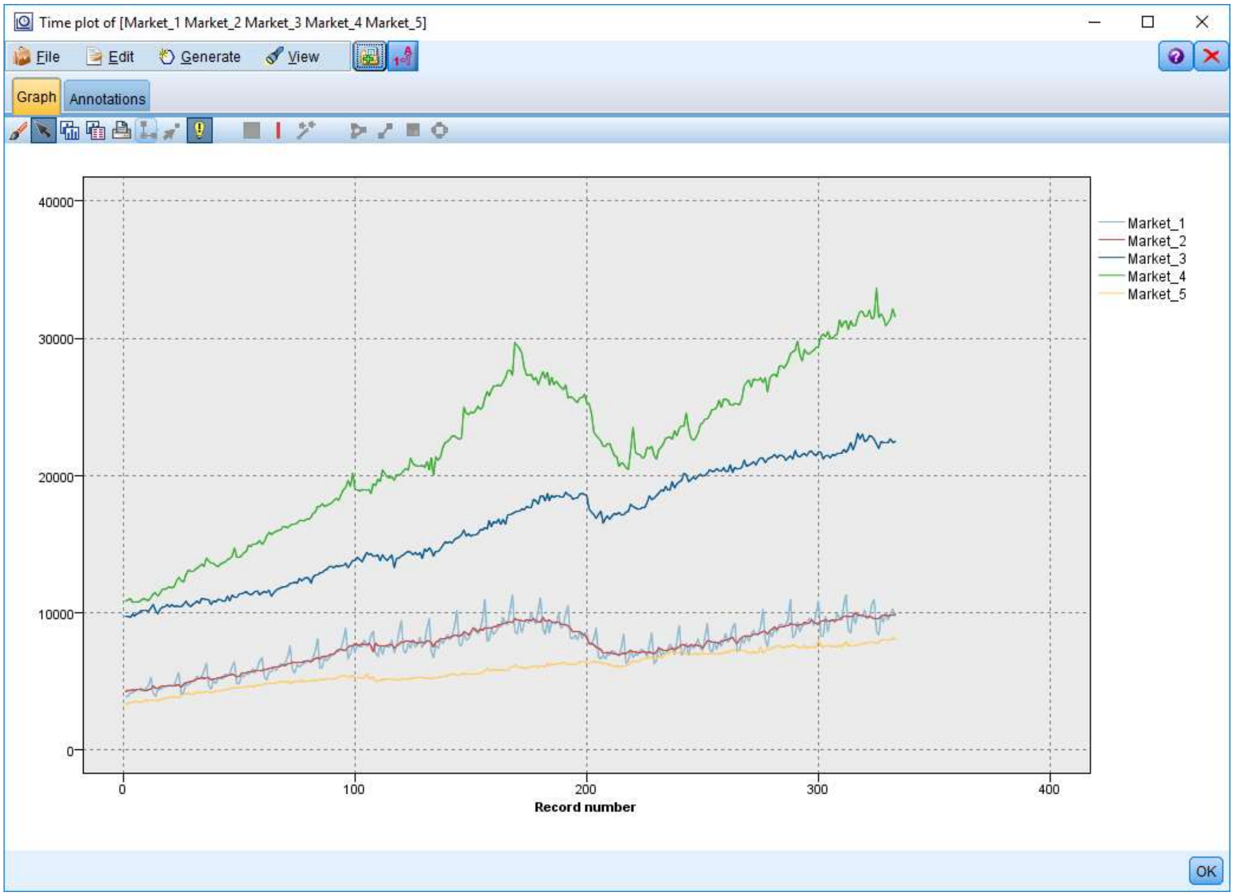
https://www.kaggle.com/census/retail-and-retailers-sales-time-series-collection#MRTSSM44111USN.csv

Process Name	Tools Used	Estimate
Data Wrangling	Excel	10 min
Modeling & Forecast	IBM SPSS Modeler	15 min
Documentation	Markdown	30 min

Wrangled Source Sample



	Market_1	Market_2	Market_3	Market_4	Market_5	Total	Date	DATE_
1	3846	4254	9716	10845	3311	319...	1-Jan-92	Jan 0092
2	3908	4342	9712	10904	3360	322...	1-Feb-92	Feb 0092
3	4157	4330	9669	10986	3445	325...	1-Mar-92	Mar 0092
4	4141	4359	9862	10738	3415	325...	1-Apr-92	Apr 0092
5	4275	4331	9755	10777	3510	326...	1-May-92	May 0092
6	4357	4401	9915	10790	3521	329...	1-Jun-92	Jun 0092
7	4407	4368	9957	10785	3470	329...	1-Jul-92	Jul 0092
8	4446	4372	10141	10915	3461	333...	1-Aug-92	Aug 0092
9	4328	4345	10154	10980	3521	333...	1-Sep-92	Sep 0092
10	4497	4299	10174	10971	3576	335...	1-Oct-92	Oct 0092

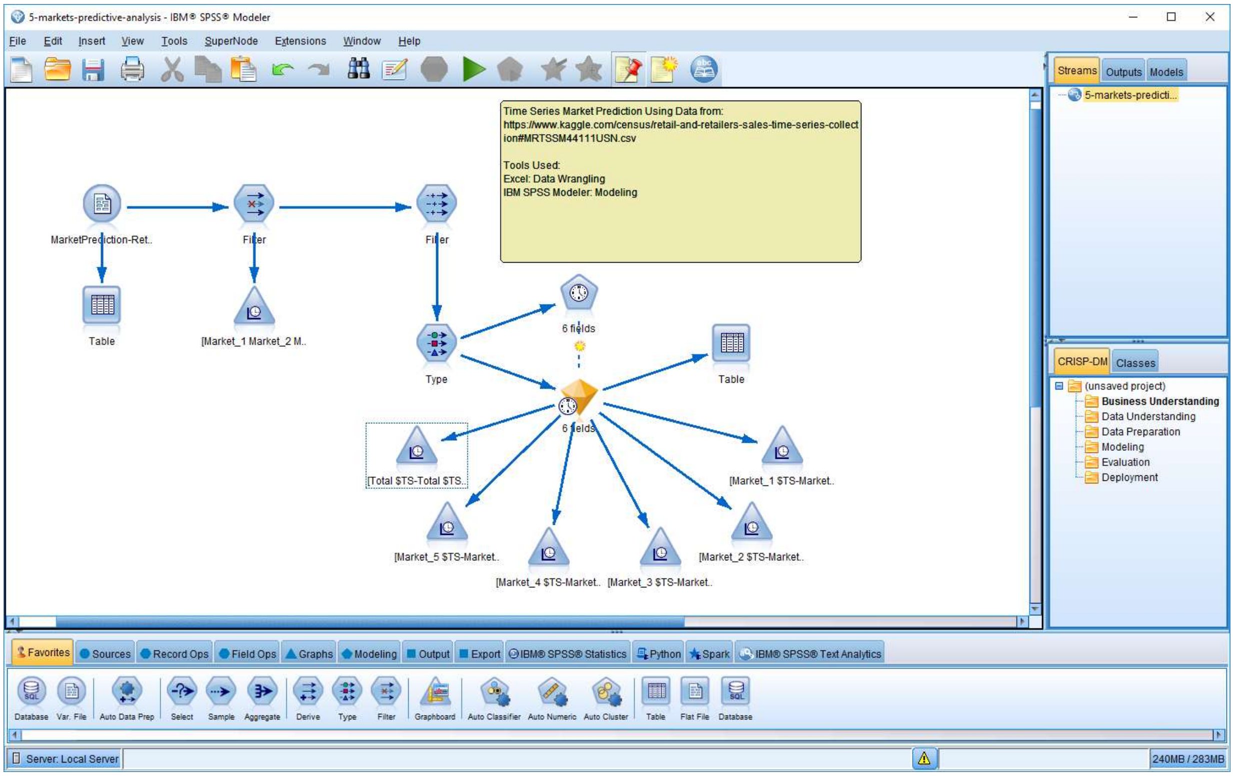


Time Series Predictive Modeling Parameters

Parameter	Setting
Time Interval	Years
Missing Value Handling	Linear Interpolation
1st forecast project interval	3
2nd forecast projection interval	200

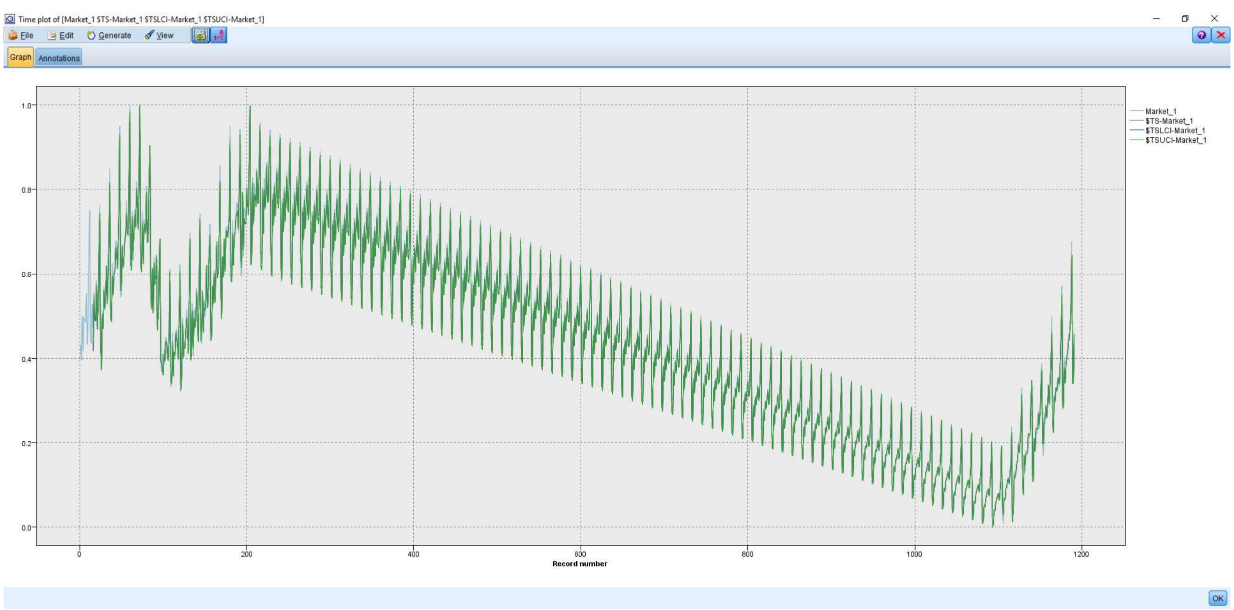
After all the parameters are set and the stream is built, I've then feed the data to train the time-series model.

Stream Design and Time Series Model Training



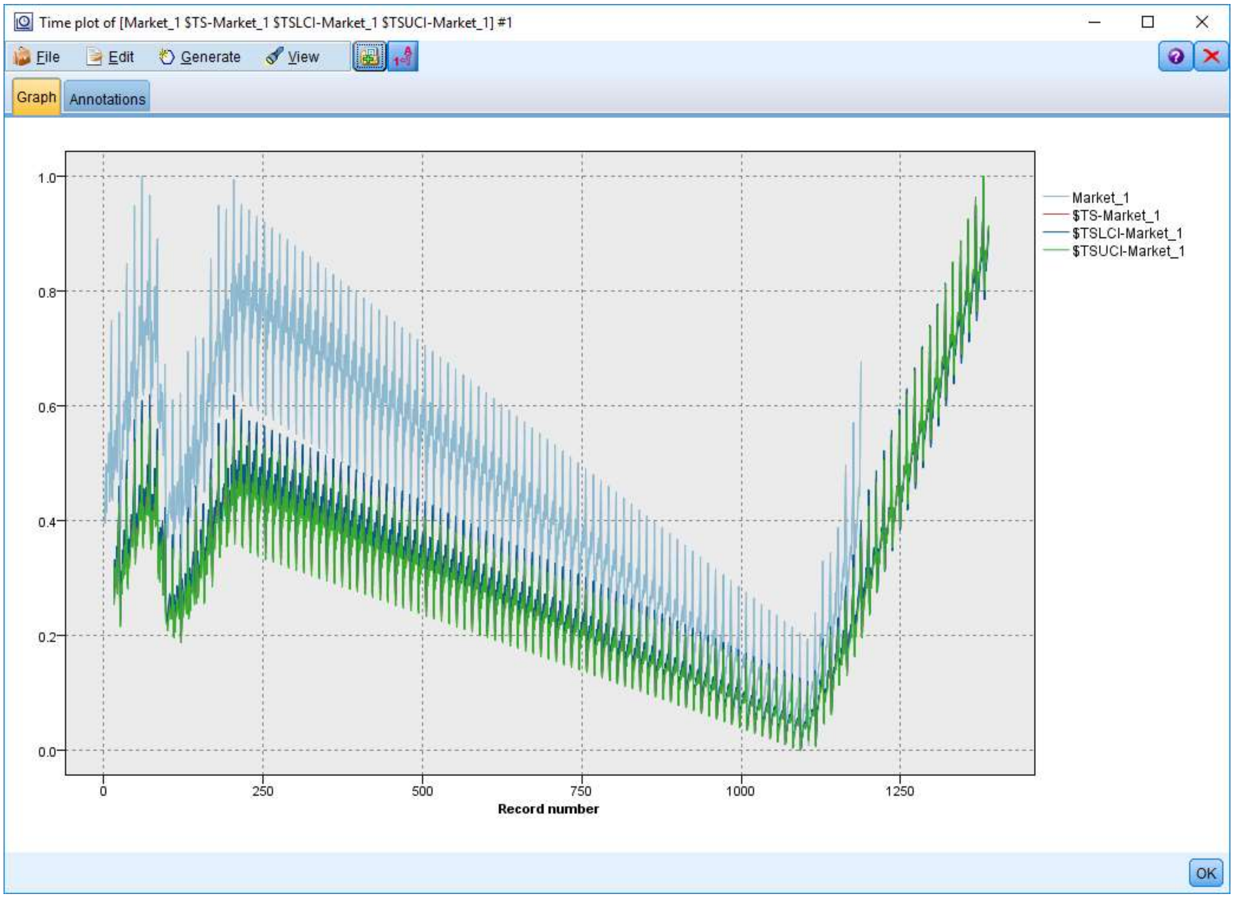
Observing Output

Benchmarking the Market_1 predictive output against real data, we can observe that the predictive model has a strong fit with the historic data.



Predicting the future Market_1 Values using the predictive model

Based on the historic data, I've computed a 200 time interval projection (5 years), we can see that the predicted value follows the market pattern versus linear trend-line, which would provides better and more accurate predictions.



Insight Discovery

In this demo project, we're interested in finding out how many predictors(i.e. Stainless Steel Market vs. Iron Ore Market) each market is using and how well do they benchmark.

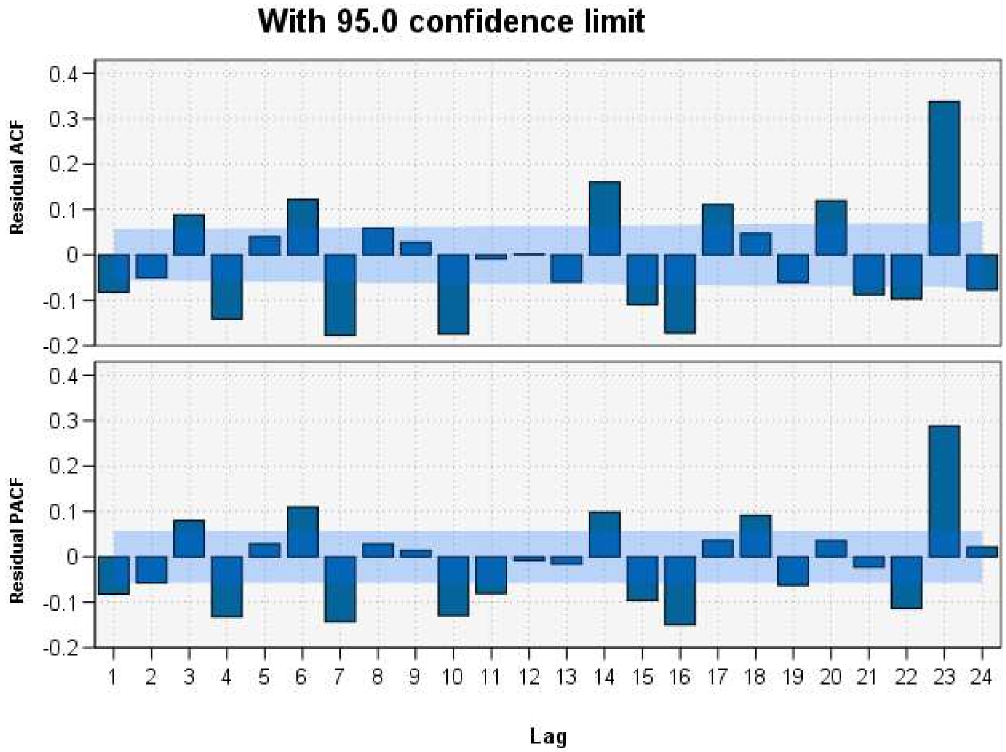
Target: Market_1

Model Information

Model Building Method		ARIMA
		Non-seasonal p=0,d=1,q=10; Seasonal p=1,d=1,q=0
Number of Predictors		2
Model Fit	MSE	3,171.835
	RMSE	56.319
	RMSPE	0.734
	MAE	24.303
	MAPE	0.324
	MAXAE	406.400
	MAXAPE	5.733
	AIC	9,468.741
	BIC	9,529.549
	R-Squared	0.999
	Stationary R-Squared	0.821
Ljung-Box Q(#)	Statistic	242.705
	df	12.0
	Significance	0.0

Parameter Estimates

				Coefficient	Std. Error	t	Significance
Market_1	No Transformation	AR, Seasonal	Lag 1	0.126	0.031	4.008	0.000
		MA	Lag 1	1.218	0.046	26.757	0.000
			Lag 3	-0.459	0.038	-12.138	0.000
			Lag 4	0.280	0.031	9.085	0.000
			Lag 9	-0.273	0.031	-8.887	0.000
			Lag 10	0.234	0.032	7.227	0.000
Market_2	No Transformation	Numerator	Lag 0	1.000	0.019	53.599	0.000
		Denominator	Lag 1	-0.073	0.024	-2.979	0.003
			Lag 2	0.096	0.018	5.200	0.000
Market_5	No Transformation	Numerator	Lag 0	-0.090	0.013	-6.870	0.000
			Lag 1	-0.086	0.012	-6.958	0.000
		Denominator	Lag 2	0.888	0.039	22.699	0.000



From the parameters estimates we can examine the Market 1 is currently using two predictors and which lag feature is used by the system as a feature / predictor. based on the Stationary R-Square, we have to say this prediction model is significantly better than Baseline Model by an index of 0.81 (maximum of Stationary R-Squared is 1)

Target: Market_4

Model Information

Model Building Method		ARIMA
		Non-seasonal p=0,d=1,q=6; Seasonal p=0,d=1,q=1
Number of Predictors		4
Model Fit	MSE	50,322.549
	RMSE	224.327
	RMSPE	0.910
	MAE	103.297
	MAPE	0.453
	MAXAE	2,095.258
	MAXAPE	6.226
	AIC	12,576.261
	BIC	12,667.272
	R-Squared	0.999
	Stationary R-Squared	0.522
Ljung-Box Q(#)	Statistic	24.796
	df	13.0
	Significance	0.0

By examining Market_4, we can find out that this market has strong connection with 4 other markets: Market 1, 2, 3, and 5, and their corresponding lag features are listed below:

This insight is important to us because in future data analysis, we can prioritize the relevant data source for more streamlined predictive analysis work.

Parameter Estimates

				Coefficient	Std. Error	t	Significance
Market_4	No Transformation	MA	Lag 1	0.262	0.029	9.115	0.000
			Lag 2	0.188	0.031	6.068	0.000
			Lag 4	0.234	0.030	7.767	0.000
			Lag 6	0.119	0.030	3.953	0.000
		MA, Seasonal	Lag 1	0.630	0.025	25.643	0.000
Market_1	No Transformation	Numerator	Lag 0	0.437	0.062	7.087	0.000
			Lag 1	-0.189	0.060	-3.128	0.002
Market_2	No Transformation	Numerator	Lag 0	0.832	0.107	7.794	0.000
			Lag 1	-0.387	0.108	-3.576	0.000
			Lag 2	-0.804	0.104	-7.745	0.000
Market_3	No Transformation	Numerator	Lag 0	0.191	0.054	3.559	0.000
			Lag 1	-0.215	0.051	-4.189	0.000
			Lag 2	0.422	0.052	8.157	0.000
		Denominator	Lag 1	0.662	0.113	5.874	0.000
			Lag 2	0.282	0.109	2.587	0.010
Market_5	No Transformation	Numerator	Lag 0	0.744	0.117	6.387	0.000
			Lag 2	-0.403	0.161	-2.503	0.012
		Denominator	Lag 1	-0.403	0.119	-3.394	0.001

With 95.0 confidence limit