||| ZS

03 : 00 : 00 : 18
DAY HRS MIN SEC

ZS Data Science Challenge - 2018

LIVE

Jul 20, 2018, 10:00 AM IST - Jul 23, 2018, 10:00 PM IST

Mekktronix Sales Forecasting Challenge

Max. Marks: 2

Problem Statement

Mekktronix is a global premium electronics manufacturing companies provides electronic devices across globe through their large distributor channel. Recently company is observing lot of fluctuations in their demand forecasting across geographies affecting their revenue. The company has reached out to build a machine learning driven forecasting solution to predict sales accurately. To test the power of AI/ML, company has provided data sales data at distributor X Week resolution as shown in Table 1.

Table 1: Format of sales dataset provided at merchant X Week level

Year	Month	Week	Merchant_ID	Product_ID	Country	Sales
2013	1	1	ar00001	1	Argentina	157500
2013	1	1	ar00003	1	Argentina	39375
2013	1	1	ar00004	1	Argentina	15750
2013	1	1	ar00007	1	Argentina	47250
2013	1	1	ar00008	1	Argentina	283500
2013	1	1	ar000011	1	Argentina	0
2013	1	1	ar000021	1	Argentina	1575
2013	1	1	ar000023	1	Argentina	47250
2013	1	1	ar000027	1	Argentina	0
2013	1	1	ar000038	1	Argentina	0
2013	1	1	ar000052	1	Argentina	277200
2013	1	1	ar000058	1	Argentina	12600
2013	1	1	ar000059	1	Argentina	126000
2013	1	1	ar000064	1	Argentina	1417500
2013	1	1	ar000065	1	Argentina	472500
2013	1	1	ar000067	1	Argentina	78750
2013	1	1	ar000068	1	Argentina	132300
2013	1	1	ar000070	1	Argentina	0
2013	1	1	ar000074	1	Argentina	94500
2013	1	1	ar000078	1	Argentina	7875

Feature Description

- **Year:** Year

- **Month:** Month

- **Week:** Week Number

- **Sales:** Price in Local Currency (Target Variable)

- **Mechant_ID :** Distributor ID

?

- **Product_Types:** Product category

- **County:** Country name

The company has requested to generate monthly forecast at Country X Product_ID resolution. The summary of training data at month level resolution is shown in table 2:

Table 2: Combination of Country X Product and Training Horizon

S. No.	Country	Product_ID	Training Period		
			From (YYYYMM)	To (YYYYMM)	Training Horizon
1	Argentina	1	201301	201603	39
2	Argentina	2	201301	201603	39
3	Argentina	3	201501	201612	24
4	Belgium	2	201301	201603	39
5	Columbia	1	201301	201603	39
6	Columbia	2	201301	201603	39
7	Columbia	3	201410	201612	28
8	Denmark	2	201301	201603	39
9	England	4	201301	201510	34
10	England	5	201301	201510	34
11	Finland	5	201301	201510	34

Feature Description

- **S. No.:** Index

- **Country:** Country with distributor dataset

- **Product_ID:** Product ID

- **From:** Start date in training data (in YYYYMM format)

- **To:** End date in training data (in YYYYMM format)

- **Training Horizon:** Number of data point in training data at monthly resolution

The company has also provided holiday and marketing expense dataset. The sample is provided in Table 3 & 4 respectively.

Table 3: Holiday dataset format

Date	Country	Holiday
2014, 1, 1	Denmark	Nytårsdag
2014, 4, 17	Denmark	Skærtorsdag
2014, 4, 18	Denmark	Langfredag
2014, 4, 20	Denmark	Påskedag
2014, 4, 21	Denmark	Anden påskedag
2014, 5, 16	Denmark	Store bededag
2014, 5, 29	Denmark	Kristi himmelfartsdag

Feature Description

- **Date:** Date in YYYY, MM, DD format

- **Country:** Country Name

- **Holiday:** Holiday detail

Table 4: Expense dataset format

Year	Month	Country	Product_Type	Expense_Price
2013	1	Argentina	3	14749.307
2013	1	Argentina	5	1329.374
2013	1	Belgium	5	249.59
2013	1	Columbia	3	1893.122
2013	1	Columbia	5	1436.726
2013	1	Denmark	5	43843.906
2013	2	Argentina	3	12187.566

Feature Description

- **Year:** year of marketing expense

- **Month:** Month of marketing expense

- **Country:** Country

- **Product_type:** Product ID

- **Expense_Price:** Expense in local currency

The client has asked for forecast at Country X Product X Month resolution. The forecasting horizon varied from {3, 9, 12} months based on country and product as shown in Figure 5. For example, for Denmark with product_ID 2 forecasting horizon is 12 months with monthly sales to be reported from Apr 2016 to March 2017.

Table 5: Forecasting summary for test dataset

S. No.	Country	Product_ID	Forecasting Period		Forecast Horizon
			From (YYYYMM)	To (YYYYMM)	
1	Argentina	1	201604	201703	12
2	Argentina	2	201604	201703	12
3	Argentina	3	201701	201703	3
4	Belgium	2	201604	201703	12
5	Columbia	1	201604	201703	12
6	Columbia	2	201604	201703	12
7	Columbia	3	201701	201703	3
8	Denmark	2	201604	201703	12
9	England	4	201511	201607	9
10	England	5	201511	201607	9
11	Finland	4	201511	201607	9

The companies have also provided an additional dataset of estimated expense at country x product_ID resolution for majority of the countries. The summary of expense details is provided in Table 6.

Table 6: Forecasting summary for test dataset

S. No.	Country	Product_ID	Expense Data Summary		Expense Data
			From (YYYYMM)	To (YYYYMM)	
1	Argentina	1	201301	201703	Available
2	Argentina	2	201301	201703	Available
3	Argentina	3	201501	201703	**Not Available**
4	Belgium	2	201301	201703	Available
5	Columbia	1	201301	201703	Available
6	Columbia	2	201301	201703	Available
7	Columbia	3	201410	201703	**Not Available**
8	Denmark	2	201301	201703	Available
9	England	4	201301	201607	Available
10	England	5	201301	201607	Available
11	Finland	4	201301	201607	Available

The final forecast must be provided at monthly resolution for Country X Product level. sMAPE will be used to evaluate the submissions.

Download Dataset here.

Evaluation

The evaluation matrix for this competition is sMAPE (Symmetric Mean Absolute Percentage Error). SMAPE is calculated as:

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^{n} \frac{|F_t - A_t|}{(|A_t| + |F_t|)/2}$$

The leaderboard score is based on SMAPE. Lower the SMAPE, higher is your score on the leaderboard.

Submission Details

Upload yds_submission2018.csv file in following format:

Year	Month	Product_ID	Country	Sales
2016	4	1	Argentina	
2016	5	1	Argentina	
2016	6	1	Argentina	
2016	7	1	Argentina	
2016	8	1	Argentina	
2016	9	1	Argentina	
2016	10	1	Argentina	

* Please follow submission guidelines for details.

** Upload the .csv in the Upload Prediction File section. While submit all other files mentioned in the submission guidelines in a .zip format in the Upload Source File section.

Starter Script

- Start with R - Click Here

Upload Prediction File
Please upload the prediction file in the format as stated in the problem.

Choose File   No file chosen

Submit & Evaluate

Upload Source Files

You need to submit a zip or tar archive consisting of a text file explaining your approach, details about feature engineering, tools you used and the relevant source files.

Choose File | No file chosen

Upload

About Us	Innovation Management	Technical Recruitment
University Program	Developers Wiki	Blog
Press	Careers	Reach Us

Site Language: English ▼ | Terms and Conditions | Privacy | © 2018 HackerEarth