

Bakery Branch Sales Prediction

Daily sales from a bakery over a time span of about 6 years and corresponding weather data for the same zone.

Possible tasks:

- Prediction of future sales values
- Prediction of future sales values with and without weather data + analysis of why and if this matters
- Prediction of weather data from sales
- Clustering of the sales data using unsupervised learning?
(try this to your own risk)

Dataset Description

We have 3 files.

- sales.csv, which contains date, group of the product, and the value of the sale for that day
- weather.csv, which contains the weather data: date, cloudiness, temperature, wind speed and weather code,
- kiwo.csv, which contains the dates for the kieler week

the files are in the ‘data’ folder.

Sales Data

```
In [3]: sales_df = pd.read_csv(os.path.join(DATA_FOLDER, SALES_DATA))  
sales_df
```

Out[3]:

	Date	Group	Sales
0	2013-07-01	1	148.828353
1	2013-07-02	1	159.793757
2	2013-07-03	1	111.885594
3	2013-07-04	1	168.864941
4	2013-07-05	1	171.280754
...
10864	2018-12-22	6	66.737353
10865	2018-12-23	6	49.958196
10866	2018-12-24	6	46.130749
10867	2018-12-27	6	51.623140
10868	2018-12-28	6	35.220810

10869 rows × 3 columns

You can use pandas to read the file (see notebook), and you see the 3 columns: Date, Group and Sales

Sales Data

- Date: the day in the format: YYYY-MM-DD
- Group: the product groups, which are

1 = Bread	2 = Small Bread (brötchen)	3 = Croissant
4 = Pastry	5 = Cakes	6 = Seasonal Breads

- Sales: the sales value (which is obfuscated, so no unit)

Weather Data

```
In [9]: weather_df = pd.read_csv(os.path.join(DATA_FOLDER, WEATHER_DATA))  
weather_df
```

Out[9]:

	Date	Cloudiness	Temperature	Wind Speed	Weather Code
0	2012-01-01	8.0	9.8250	14	58.0
1	2012-01-02	7.0	7.4375	12	NaN
2	2012-01-03	8.0	5.5375	18	63.0
3	2012-01-04	4.0	5.6875	19	80.0
4	2012-01-05	6.0	5.3000	23	80.0
...
2596	2019-07-28	3.0	23.3500	14	5.0
2597	2019-07-29	6.0	25.2500	7	61.0
2598	2019-07-30	7.0	20.7375	8	61.0
2599	2019-07-31	6.0	20.4500	7	61.0
2600	2019-08-01	5.0	21.0625	9	61.0

2601 rows × 5 columns

Columns: Date, Cloudiness, Temperature, Wind Speed and Weather Code

Weather Data

- Date: the day in the format: YYYY-MM-DD
- Cloudiness: daily average, from 0 (min) to 8 (max)
- Temperature: daily average temperature in Celsius
- Wind Speed: daily average speed in m/s
- Weather Code: the (german) legend is here:
http://www.seewetter-kiel.de/seewetter/daten_symbole.htm
In case your group do not understand german, please ask for help

Real Life Scenario

meteolytix forecast analysiert die Datenzusammenhänge von mehr als 400 Einflussfaktoren und liefert Absatzprognosen für viele Einsatzfelder.

WAS WIR MACHEN

