# UDatE

## Data Quality

1. Data description: Daily sales for 1,115 drug stores located across Germany.
2. Data shape:

* Store: 1,115 rows × 10 columns Means factor variables for each 1,115 drug store.

Features: *Store type; assortment*; *CompetitionDistance*; *CompetitionOpenSinceMonth*; *CompetitionOpenSinceYear*; *Promo2*; *Promo2SinceWeek*; *Promo2SinceYear*; *PromoInterval*

表格

描述已自动生成

表格

描述已自动生成

*CompetitionDisctance*: from 20 to 75860

* Train: 1017209 rows × 8 columns, by date, from 01/01/2013 to 31/07/2015; each store has 942 records.

Features: *Store*; *DayOfWeek*; *Date*; *Open*; *Promo*; *StateHoliday*; *SchoolHoliday*

Output/Y value: *Sales* & *Customers*表格

描述已自动生成

* Test: Predict from 01/08/2015 to 17/09/2015

1. Variables

●  Time trend

●  Past sales

●  StoreType

●  Assortment

●  Competition

●  Promo2

●  DayOfWeek

**●  Open**

●  Promo

●  Holiday

1. Missing value

* Store had two types of missing values:

1. Competition: *CompetitionOpenSinceMonth* & *CompetitionOpenSinceYear*
2. If we look at the *CompetitionOpenSinceYear*, we can find 570 stores’ competitors opened before 2013(the year started to record data).
3. If we compare the distance with missing value, take the median 2325 km as a boundary point, we can get there are 167 stores has competitors shorter than 2325 km. The distance doesn’t affect the missing value.
4. There’s no row that only has *CompetitionOpenSinceMonth* value or *CompetitionOpenSinceYear* value.
5. Promotion: Promo2SinceWeek & Promo2SinceYear & PromoInterval

Here the missing value means the store doesn’t participate the Promo2.

文本, 信件

描述已自动生成

1. Hypothesis

* Would competitor affect stores’ sales?

Question here: The appearance of competitor is in a sudden and have a permanent effect, especially for those who have competitors earlier than 2013 year, should we still take consider the competitor into account?

* Would promotions affect stores’ sales?
* Would holiday affect stores’ sales?
* Would past 7 days sales data affect current sales?