CARD

ID

Sentence describing the usage scenario

[The Rossmann sales director wants to have a system for forecasting, up to 6 weeks, the daily sales for each of the 1115 stores. Store sales are influenced by many factors, including promotions, competition, school and state holidays, seasonality, and locality. The forecasts are necessary for the staffing purpose.]

Acceptance test

[The sales director wants to reach an out-of-sample accuracy of 60% for the smaller store and 80% for bigger store, where the prediction error cause more problems]

Time [Hours/Days]: Developers responsibility

[10 Days]

Value: Product Owner responsibility

[ 5/10 ]

Tool for staff – charts – visualize data – graphical results –

Table –

Predictive –

Data extracting – to query local databases – retrieve data to feed the existent database. For the POC I can use Kaggle dataset .

Querying external databases –

Background: 10 lines, data I want to analyse,

How the sw solution the problem of the customer.

Scope: end result –

Proof-of-Concepts: 3-4 functionalities (web scraping for the future – connect with marketing database)

Economical offer: first release – cost? next release: - additionally cost. For the next year next phase

Goals: goal, target for sales director… measurable, can be tested by your customer

Customizable chart, window – select 1,2 parameter …

product:

• Data visualization tool to expose trends

• Data forecaster

Requirements

a) The sales director would access a dashboard with a few visualizations identifying some interesting

trends for the sales in the different stores and areas (aggregate level) and the TOP-N stores.

b) The store director should see just the data for its store and store of the same category but thery cannot see the sales of all other stores.

Domain

Dataset provided by the company. Additionally, the developer can make use of “external” data (consumer confidence index, inflation, GDP growth …) to try make better forecasting.

deliverables

1. project charter (formal documents, initial presentation, include the goal of the project and data needed: product & methodology ) 
2. Proof of Concept – not to trivial and not too complicated!
3. full-fledged project Plan

# 1 project charter

### Background

The idea of this project is to realize a tool for the sales forecasting. The sponsor of this project is the sales director of Rossman (the iconic German drugstore chain) and the tool will be used by the marketing departments that needs a more sophisticated statistical analysis than the actual BI tool (Microsoft Dynamics). This is a pilot project customized for the German stores, but it could be easily adapted to other location area.

The tool should extract from the DWH the raw data (sales for each day in each department, active promotions) and should enrich it with additional information external regarding such as

* holidays (state holidays, school holidays)
* weather data
* macroeconomics indicators like:
  + GfK Consumer Confidence is a leading index that measures the level of consumer confidence in economic activity
  + Consumer confidence index (CCI)
  + …

The tool should be identifying the characteristics of the time series (presence of trend, seasonality, serial correlation, heteroscedasticity, breaks points) and should be able to forecast the sales.

The forecasts are necessary for budgeting purposes (PL forecast) but could be used for other purposes: staffing, logistics.

### Scope

Scope of this pilot project is limited to the statistical analysis of daily sales of all 1115 Rossman store for the period “2013-01-01” to “2015-07-31” and the realization of a POC with use of Power BI and R-python code that will able to produce:

* graphical visualizations (up to 6 Power BI slides)
* the forecast matrix (daily forecast for each store for the following 6 weeks).

For the statical analysis will be considered only the internal data and the macroeconomics indicators. Weather data won’t be considered since web scraping is needed and will be object of a possible extension of this project.

The necessary data preparation analysis phase will be performed “one off” (for the purpose of the POC, using simple manual tools) and will be not provided. Additional activities for the data preparation to be used for the on-going process (such as data extraction, integration, data reconciliation with the legacy system) are out of scope.

### Goals

• dataset preparation / selection

• identification of the characteristics of the time series

• model selection

• POC tool selection

• implementation of POC

• tests with different series and model calibration

• test with the final users

### Future steps:

• data extraction/ integration for extracting the data from the DWH



• realization of a specific DataMart

• web scraping for the weather data

• web scraping for the external macroeconomic data

### Metrics

•

• What are the qualitative objectives? (e.g. reduce user churn)

• Define a quantifiable metric , and quantify what improvement in the values of the metrics are

useful for the customer scenario (e.g. reduce the fraction of users with 4-week inactivity by 20%) 

MOV. What is the baseline (current) value of the metric? (e.g. current fraction of users with 4-week

inactivity = 60%) How will we measure the metric? (e.g. A/B test on a specified subset for a

specified period; or comparison of performance after implementation to baseline)

### Personnel involved

• Who are on this project:

o Us: Project lead, PM, Data scientist(s), Developers, Account manager, Sysadm…..

o The Client: Data administrator, Business contact, CTO,

### Key Stakeholders (other than personnel)

Client [name]

Sponsor [name]

Project manager [name]

Project team members [name], [name], [name], [name].

### Project Milestones

[Identify the S.M.A.R.T. project milestones: start date, end date and invoicing dates to the client.]

### Project Budget

[Describe the main project expenses: non-recurring & monthly recurring.]

### Data Architecture

• What data do we expect? Raw data in the customer data sources (e.g. on-prem files, SQL, on-prem

Hadoop etc.)

• What tools and data storage/analytics resources will be used in the solution?

• How will the customer use the model results to make decisions?

• architecture before and after the project

• End to end data flow

• Data movement pipeline in production

### Constraints, Assumptions, Risks and Dependencies

Constraints [Describe here potential factors that will impact the delivery of the project]

Assumptions [Describe here conditions or situations that you are relying on in order to achieve

project goals, ex. Frequency of meetings, data availability, customers’ resources

availability etc.)]

Risks and Dependencies [What are the most significant risks? What things must happen before the

project is delivered?]

Approval Signatures

[Name], Project Client [Name], Project Sponsor [Name], Project Manager

# 2 Proof of Concept

# 3 Full-fledged project Plan