ROSSMAN STORE SALES FORECAST & APP - PROPOSAL PRESENTATION

FROM TEAM ASTUTE IT SOLUTIONS INC.,

BY

HENRY BERNREUTER

YUN LIU

SINDHU VATTIVELLA

AGENDA

- About the Company
- Primary Objective
- Problem Description
- Solution Prominence
- Evaluation Criteria
- Application Development Plan
- Data Sourcing
- Data Integration
- Data Storage
- Features
- Challenges
- Future Aspects

ABOUT THE COMPANY

- Rossman is Germany's second-largest drug store chain , with over 3,790 stores in Europe.
- Rossmann was founded in 1972 by Dirk Rossmann
- It Headquartered in German town
  of Burgwedel near Hanover

OBJECTIVE

- To forecast the daily sale of individual 1115 Rossmann stores located across Germany, 6 weeks in advance.
- The Sales forecast is done by creating a robust sales prediction model
- To Create an

PROBLEM DESCRIPTION

Business Problem

- The constant battle to capture, trap and convert visits into sales revenue had led companies to seek new solutions for streaming real time data to make real time sales solutions.
- The problem is to fuel the company’s growth by applying a predictive model to all pricing solutions.

SOLUTION PROMINENCE

- Optimize staff schedules.
- Provide more time to store managers to keep their focus on customers
- Increases efficiency of employees.

EVALUATION CRITERIA

- The Solutions are evaluated on the Root Mean Square Percentage Error(RMSPE).
- Lower the score better will be the prediction.

$$RMSPE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - \hat{y}_i}{y_i}\right)^2}$$

- y(i) - Sales of a single store on a single day
- y(i) _hat - Corresponding prediction.
- Anyday, store with 0 sales is excluded in scoring.

Architecture

Below is the architecture designed for our application. The Initial phase consists of the below components for the application.

APPLICATION DEVELOPMENT PLAN

DATA SOURCING

- Dataset consists of Historical sales data for 1,115 Rossmann stores.
- It also comprises of promotional and competitor data.
- Some of the main data definitions are
  - Store Info
  - Holiday Info
  - Sales
  - Promos Info
- Sales column must be predicted.

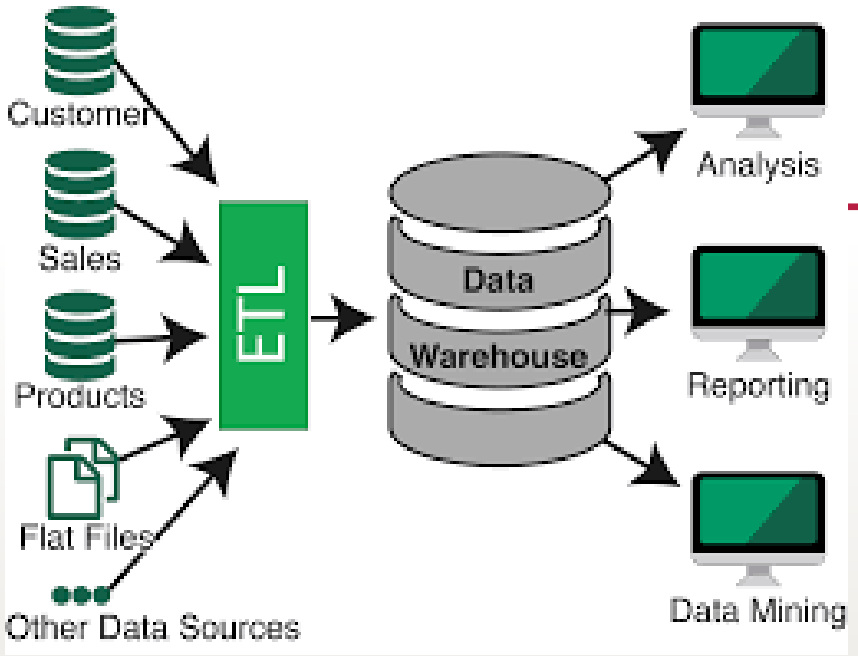
Dataset statistics

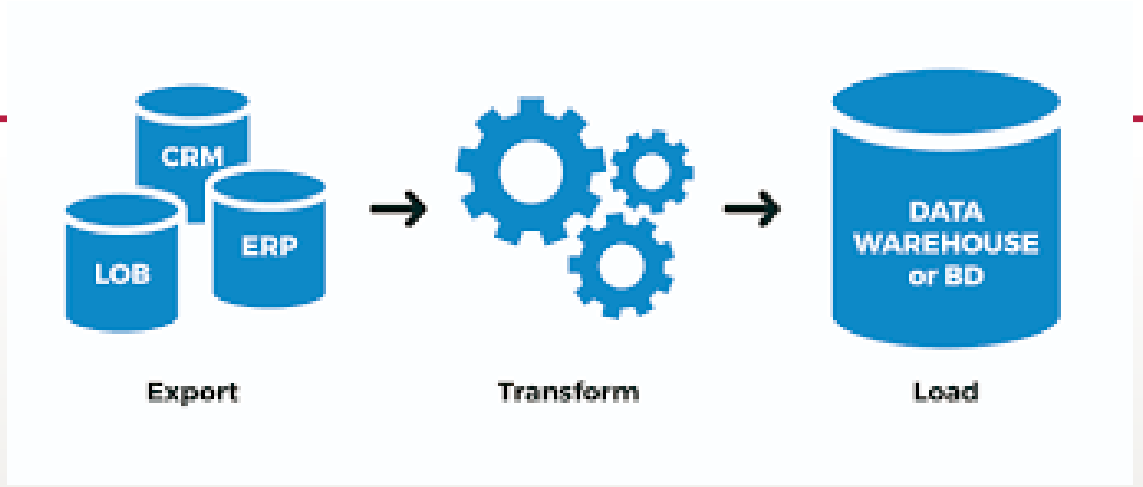
STATISTICS	NUMBERS
Dataset size	1017209
Testing data size	41088
Total stores number	1115
Training data Time ranges	2013-01-01 to 2015-07-31
Testing data Time ranges	2015-08-01 to 2015-09-17

Table 2

Sales statistics

STATISTICS	VALUES
Global store sales average	5773.82
Max daily sales	41551
Min daily sales	46



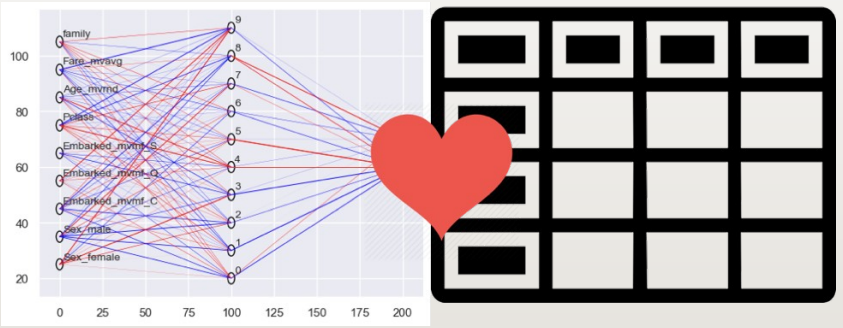


DATA INTEGRATION

- Extracting information
- Transforming the information
- Loading the information in structured format for further analysis.

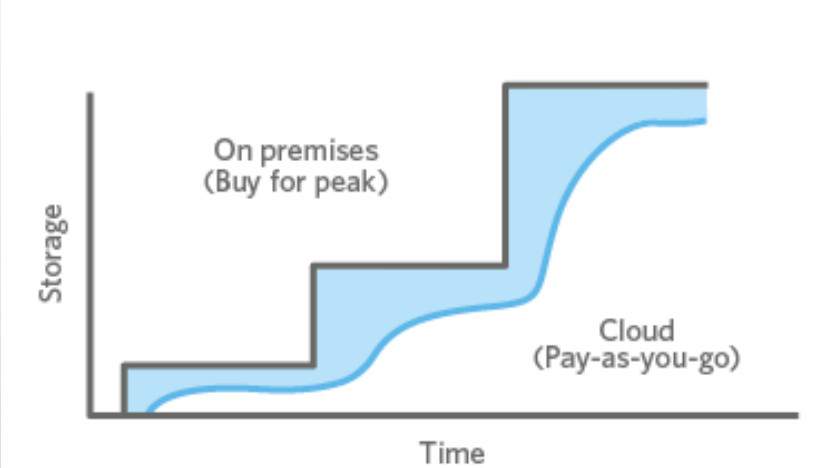
DEEP NEURAL NETS

- Effective technique for tabular data analysis, requiring little feature engineering and less maintenance than other techniques.
- Reliable and effective for making predictions on tabular data.
- There are many tabular data analysis tasks that a deep neural network model can be trained to perform:
  - Fraud detection
  - Sales forecasting
  - Product failure prediction
  - Pricing
  - Credit risk
  - Customer retention/churn



DATA STORAGE

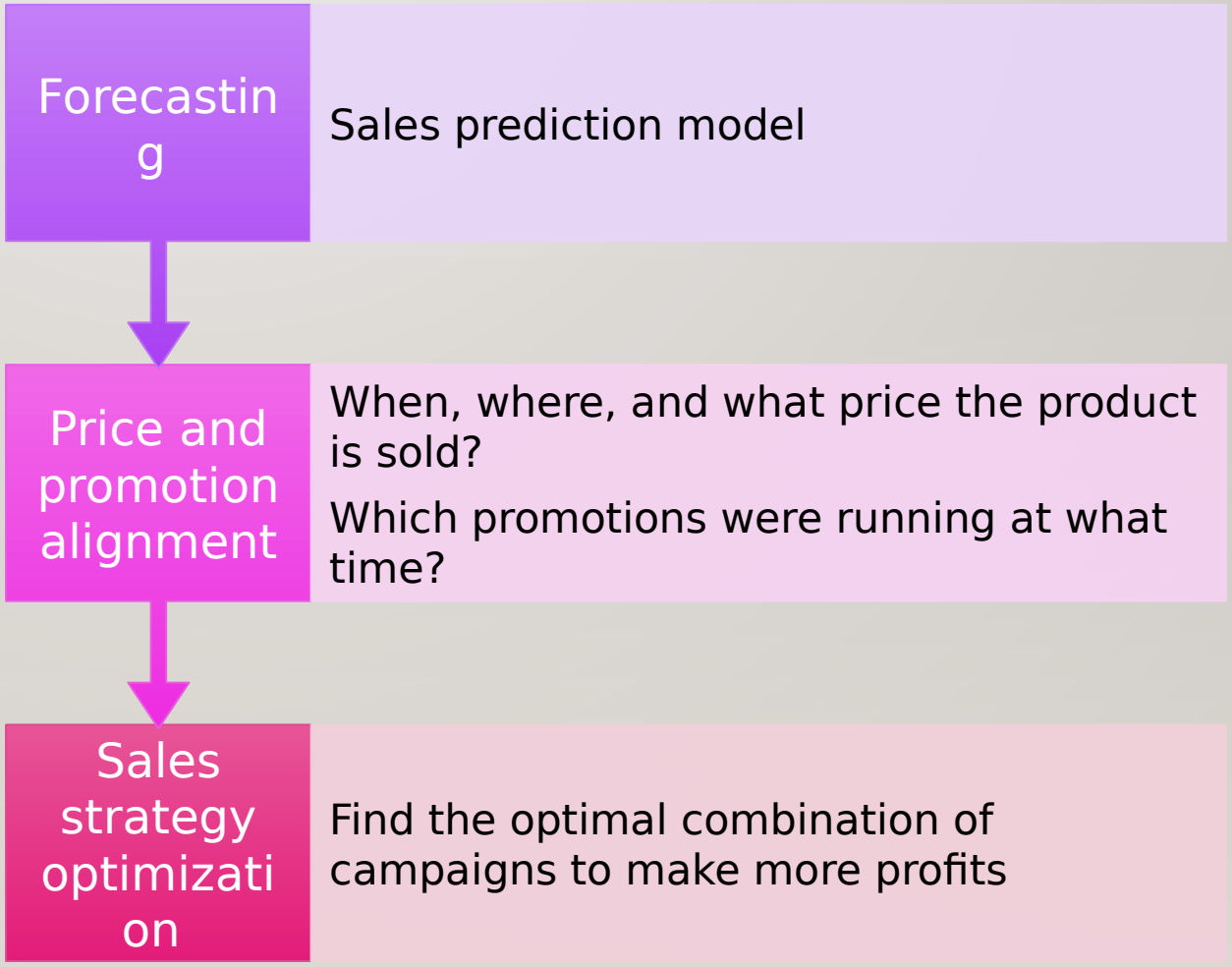
- Data resides in Aws cloud environment I.e., Specifically in the form of s3 buckets.
- Amazon S3 provides the highest level of data durability and availability on the **AWS** Cloud.
- Gateway uploads data from the upload buffer over an encrypted Secure Sockets Layer (SSL) connection to **AWS Storage** Gateway service(in **AWS** Cloud).
- The service then stores the data encrypted in Amazon **S3**.
- One can take incremental backups, called snapshots, of **storage** volumes.



DATA RETRIEVAL

- Data is retrieved from S3 buckets into R environment.
- It can be done by Connecting AWS S3 to R via., aws.s3 package.
- Three main steps in data retrieval process are
  - Setting up credentials to connect R to S3
  - Authenticating with aws.s3(Package installation and Token's specification)
  - Reading and writing data from/to S3

FEATURES


Forecasting
Sales prediction model
Price and promotion alignment
When, where, and what price the product is sold?
Which promotions were running at what time?
Sales strategy optimization
Find the optimal combination of campaigns to make more profits

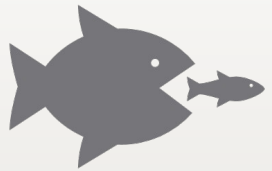
CHALLENGES

- Incompleteness
- Promotional cannibalization
- Product specificity and seasonality
- New product

FUTURE ASPECTS OF SALES FORECASTING

High accuracy

Widely used in the retail industry



Using for competition

REFERENCES

- HTTPS://MEDIUM.COM/HASHMAPINC/ETL-UNDERSTANDING-IT-AND-EFFECTIVELY-USING-IT-F827A5B3E54D
- HTTPS://WWW.SCIENCEDIRECT.COM/SCIENCE/ARTICLE/ABS/PII/S0925231207001622?VIA%3DIHUB
- HTTPS://WWW.TECHOPEDIA.COM/DEFINITION/30140/DATA-RETRIEVAL
- HTTPS://IOPSCIENCE.IOP.ORG/ARTICLE/10.1088/1742-6596/1192/1/012010/META

THANK YOU