Data Mining Approach to Predict Campaign Result

With Bank Marketing Dataset

Chunyi Bi

Project Background (1/3) : Bank TeleMarketing

A Portuguese retail bank uses marketing selling campaigns to enhance business. The bank uses call center to contact target customers and encourage them to subscribe a certain product.

Campaigns

Call Center
↓
Target Customer
↓
Subscribe?

Main Goal

Propose data mining approaches to predict the success of telemarketing calls for selling long-term deposits

Could we

➤ Help the bank to develop deeper understanding of the target customers and establish customer profile?
➤ Increase the effectiveness of the bank's telemarketing campaign?

Project Background (3/3) : DATASET

Dataset is collected from 2008 to 2013, containing customer related data and campaign related data.
21 Attributes, 41,188 instances, and no missing value.

| CATEGORY	#	ATTRIBUTES and DESCRIPTION		TYPE
Bank Client data	1	age	age of the customer	numeric
	2	job	type of job	categorical
	3	marital	marital status	categorical
	4	education	education level	categorical
	5	default	has credit in default?	binary
	6	housing	has housing loan?	binary
	7	loan	has personal loan?	binary
the most recent Contact Data	8	contact	contact communication type	categorical
	9	month	last contact month of year	categorical
	10	day_or_week	last contact day of week	categorical
	11	duration	last contact duration, in seconds	numeric
Campaign Data	12	campaign	number of contacts performed during this campaign	numeric
	13	pdays	number of days that passed by the most recent contact	numeric
	14	previous	number of contacts performed before this campaign	numeric
	15	poutcome	outcome of the previous marketing campaign	categorical
Social and Eco	16-20	emp.var.rate, cons.price.idx, cons.conf.idx, euribor3m, nr.employed		numeric
Target Output	21	y	has the client subscribed a term deposit?	binary

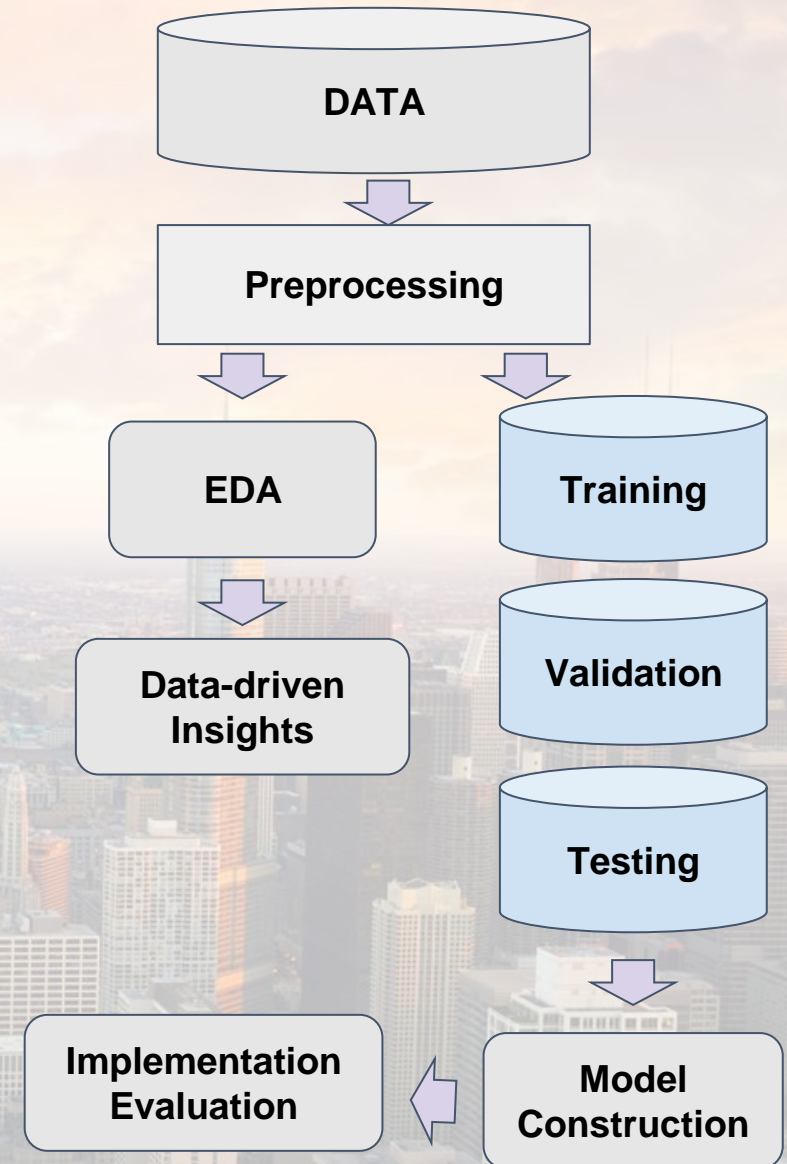# Visualization
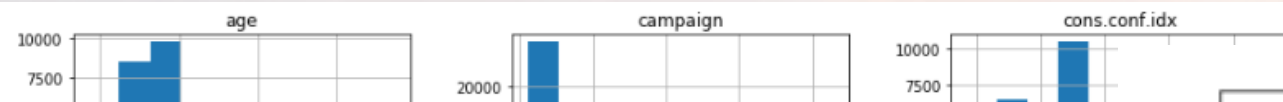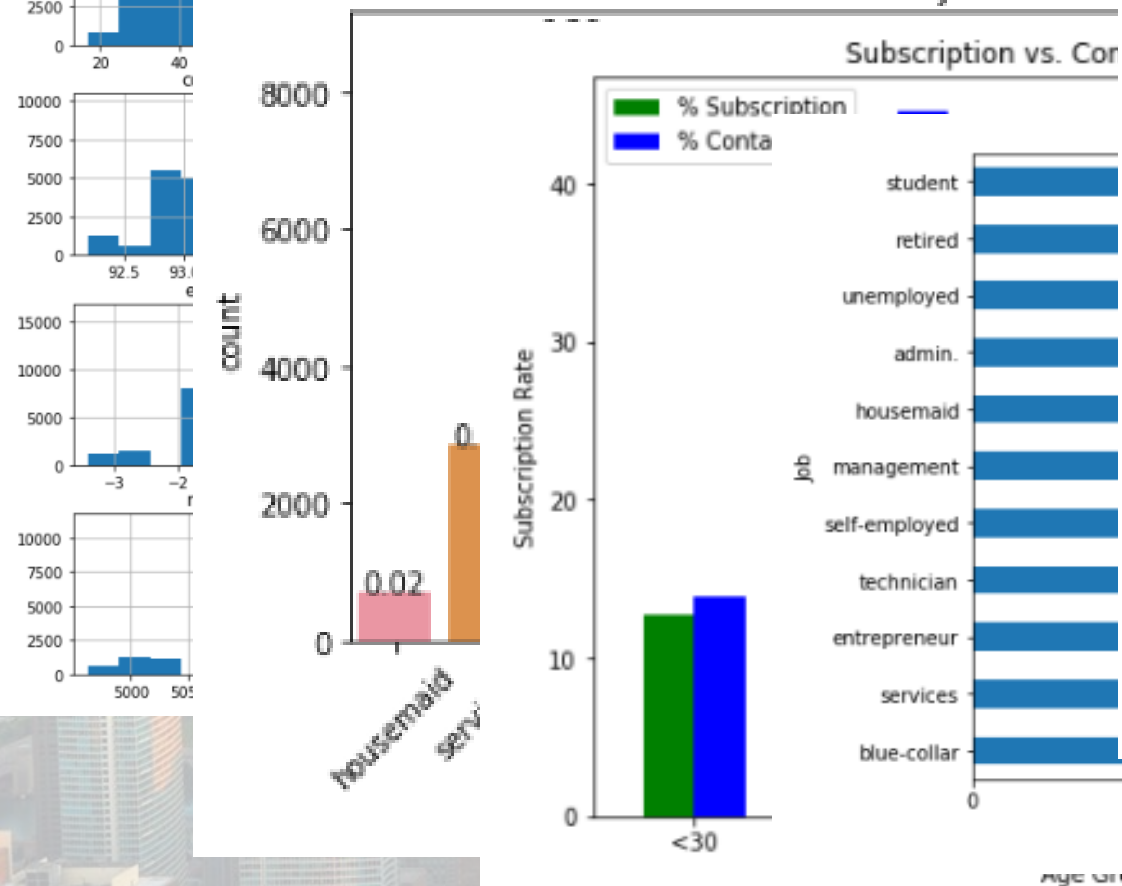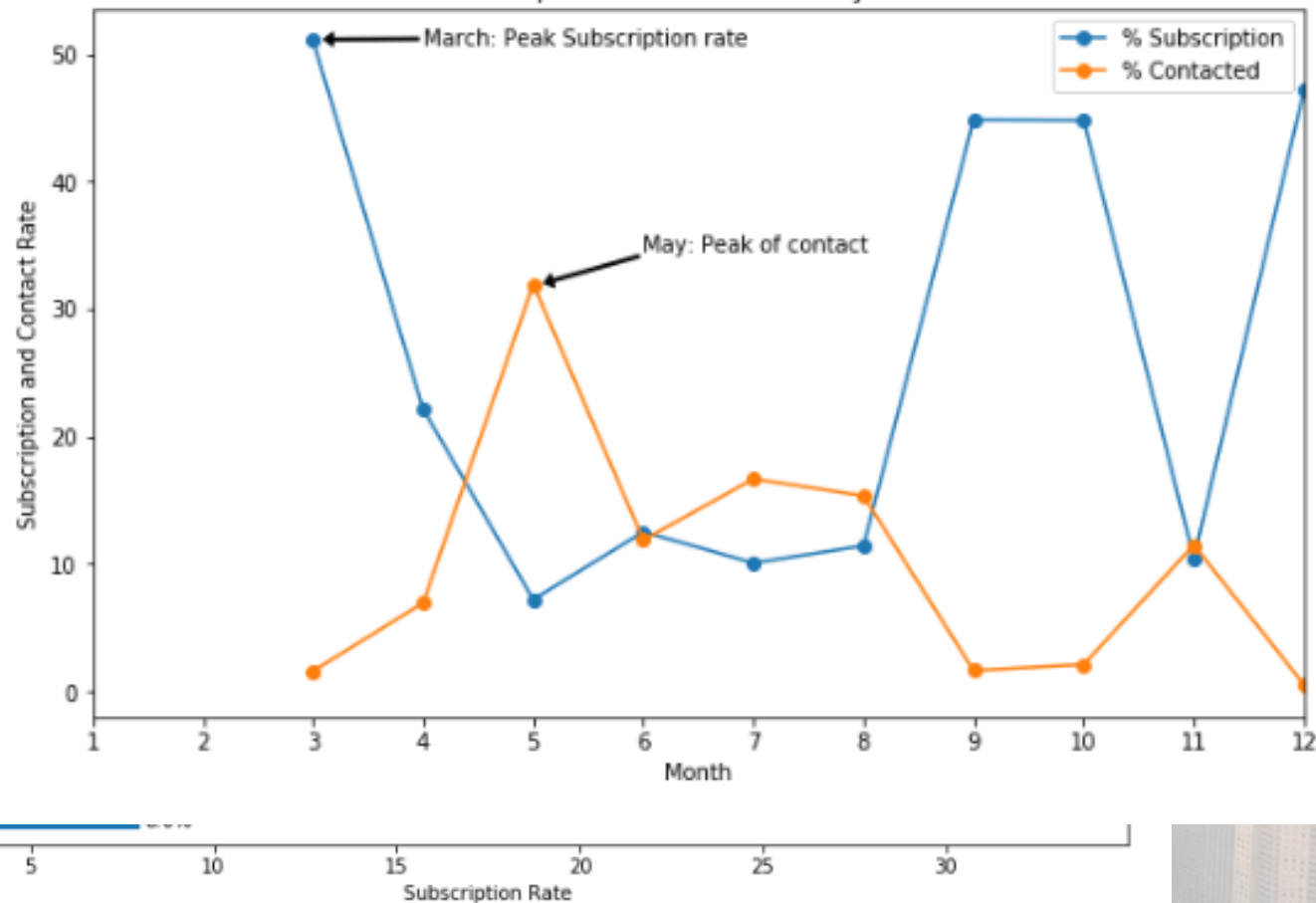
Classification Model

➢ Data Preprocessing:

  ○ missing value

  ○ duplicated rows

  ○ data transformation (one-hot encoding, label encoding)

  ○ SMOTE upsampling

  ○ Scale

➢ Model Building:

  ○ Evaluation matrix: AUC score

  ○ Baseline model: Logistic Regression

  ○ Parameter Tuning: Logistic Regression/ KNN

  ○ Final model

  ○ Implementation

DATA

Preprocessing

EDA

Data-driven Insights

Training

Validation

Testing

Implementation Evaluation

Model Construction

Managerial Conclusions

Goal: Propose data mining approaches to predict the success of telemarketing calls for selling long term deposits

Insights

➤ The bank should target the youngest and the oldest group, instead of middle-aged group.
➤ The bank should target the students and retired customer group.
➤ The bank should initiate campaign during spring and fall.
➤ The bank should pay attention to the social and economic indicators.

THANK YOU