

PREDICTING BANK TELEMARKETING SALES

Nick Gigliotti



BANCO DE
PORTUGAL
EUROSISTEMA

Business Problem

Banco de Portugal hired me to develop an accurate **predictive model** to predict which customers are likely to invest in a term deposit as a result of telemarketing.



Telemarketing is hard on both salespeople and customers.



Connecting with the wrong customers irritates them and **produces no profit.**



Connecting with high—potential customers **increases profit** and customer satisfaction.

Dataset

- Bank telemarketing dataset
- Collected by an **anonymous** Portuguese retail bank
- Publicly available on the [UCI Machine Learning Repository](#)
 - Compiled in 2014 by:
 - Sérgio Moro (ISCTE-IUL),
 - Paulo Cortez (Univ. Minho)
 - Paulo Rita (ISCTE-IUL)
- Collected from May 2008 to November 2010
 - Date is not provided
 - Samples are ordered by date (nevertheless)
- Originally contains 21 features and 41k samples
- Roughly two thirds categorical and one third numeric
- Supplemented with economic context indicators
 - Sourced from Banco de Portugal



Feature Explanations



- Client Information

- *age* — years
- *job* — type of job
- *marital* — marital status
- *education* — level of education
- *housing* — has housing loan
- *loan* — has personal loan

- Current Campaign

- *contact_cellular* — last contact was on cellular (vs. landline)
- *contact_month* — month of last contact
- *contact_weekday* — weekday of last contact
- *contact_duration* — duration of last contact in seconds
- *contact_count* — total number of contacts during this campaign
- *invested* — invested in a term deposit (target variable)

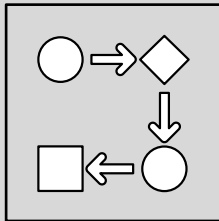
- Previous Campaigns

- *prev_contact* — contacted at least once before current campaign
- *recent_prev_contact* — 7 days or less since contact during previous campaign
- *prev_success* — previous campaign resulted in a sale
- *prev_failure* — previous campaign resulted in no sale

- Economic Context

- *emp_var_rate* — employment variation rate (quarterly indicator)
- *cons_price_idx* — consumer price index (monthly indicator)
- *cons_conf_idx* — consumer confidence index (monthly indicator)
- *euribor_3m* — [Euribor](#) 3-month rate (daily indicator)
- *n_employed* — thousands of people employed (quarterly indicator)

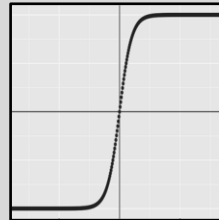
Methods and Model



Iterative Process

Logistic Regression

- Relatively simple
- Powerful
- Widely used

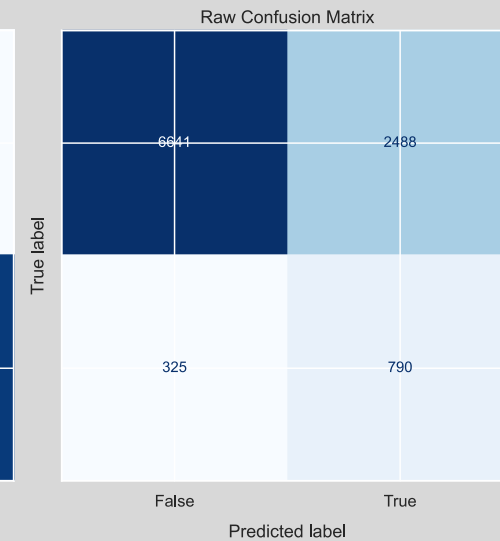
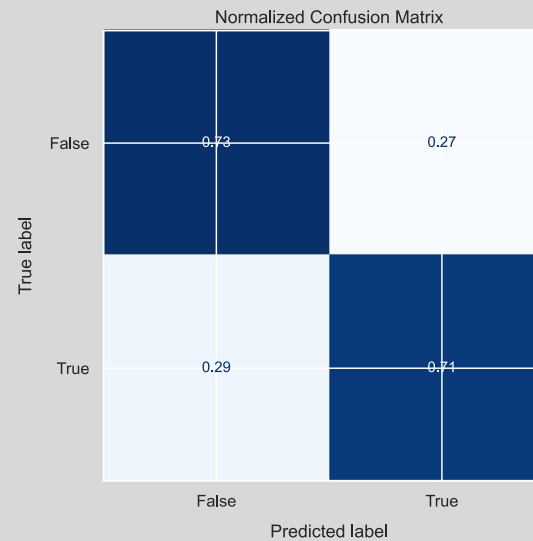
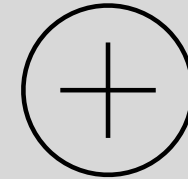


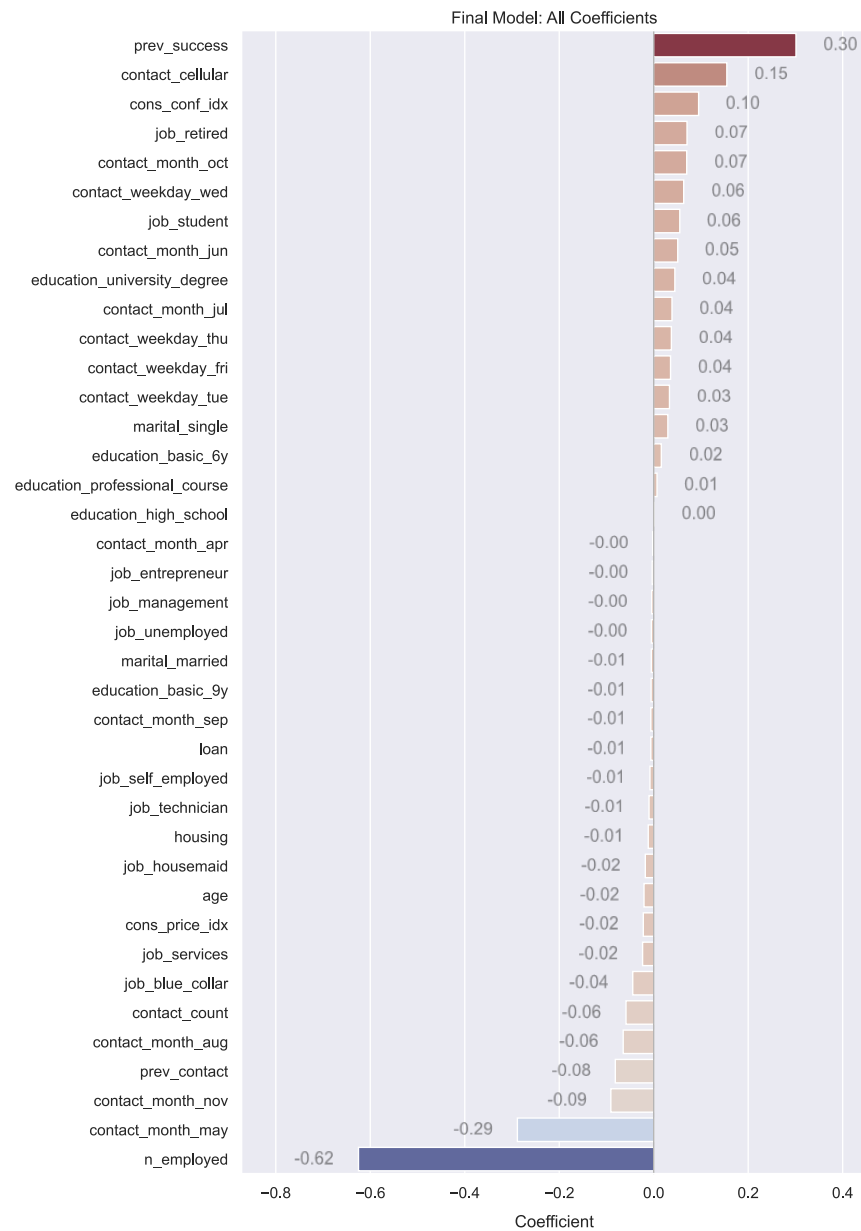
Preprocessing

- Fill missing values
- Filter out highly correlated sets of features
- Reduce the influence of outliers
- Scale the data

Final Model Evaluation

- Balanced accuracy: 0.72
 - Average of positive and negative recall
- Strong diagonal

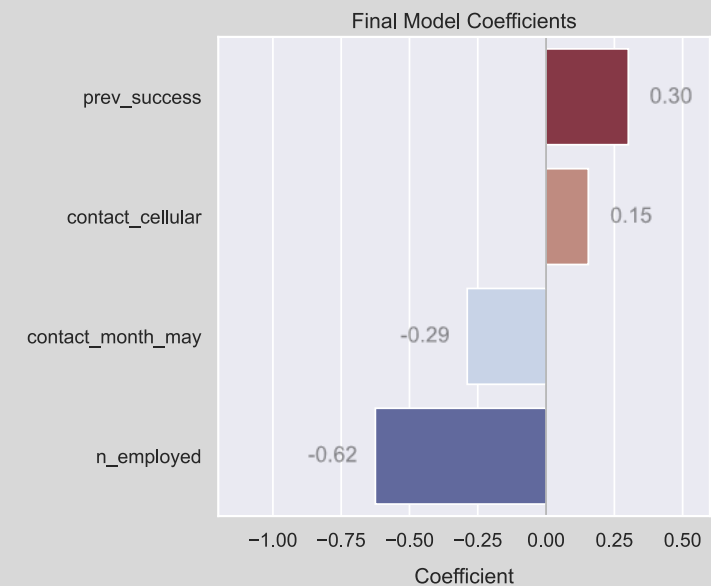




Top Features For Prediction

Ordered by magnitude:

- n_employed — Portuguese employment count
- prev_success — previous marketing campaign resulted in a sale
- contact_month_may — contacted in May
- contact_cellular — contacted on a cell phone as opposed to a landline



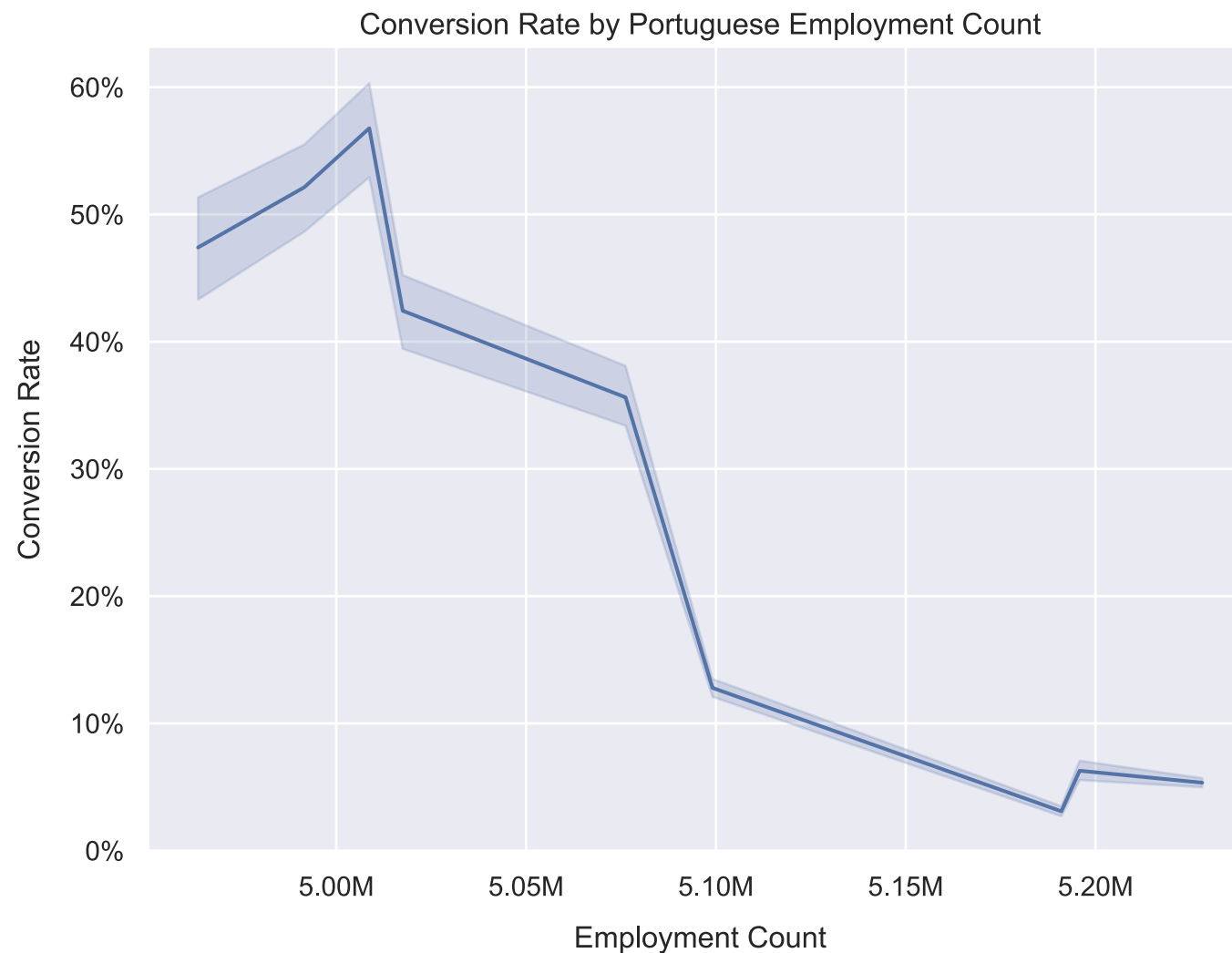
Push Hard When Employment is Low

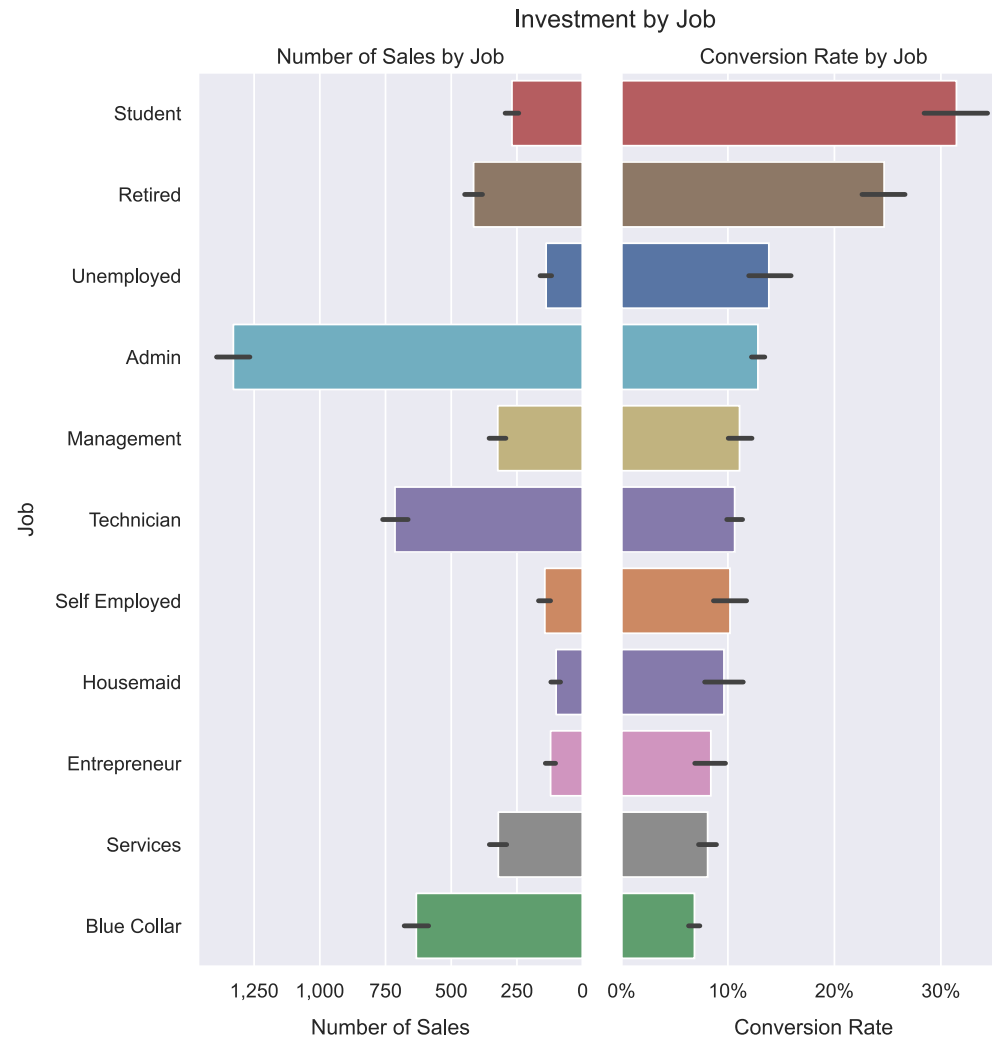
- Undoubtedly, a very strong relationship
- The underlying mechanism behind this relationship is unknown to me
- A surprising discovery resulting from my model
- What is the conversion rate for unemployed people?

Recommendation:

Pump resources into marketing when employment is **low**.

Relax your efforts when employment is **high**.





Unemployed People Don't Invest

- The conversion rate for the unemployed is surprisingly high, but not that high
- The total sales to unemployed people is predictably low
- Students and retirees have the highest average conversion rate
- Administrators and technicians have the highest total sales

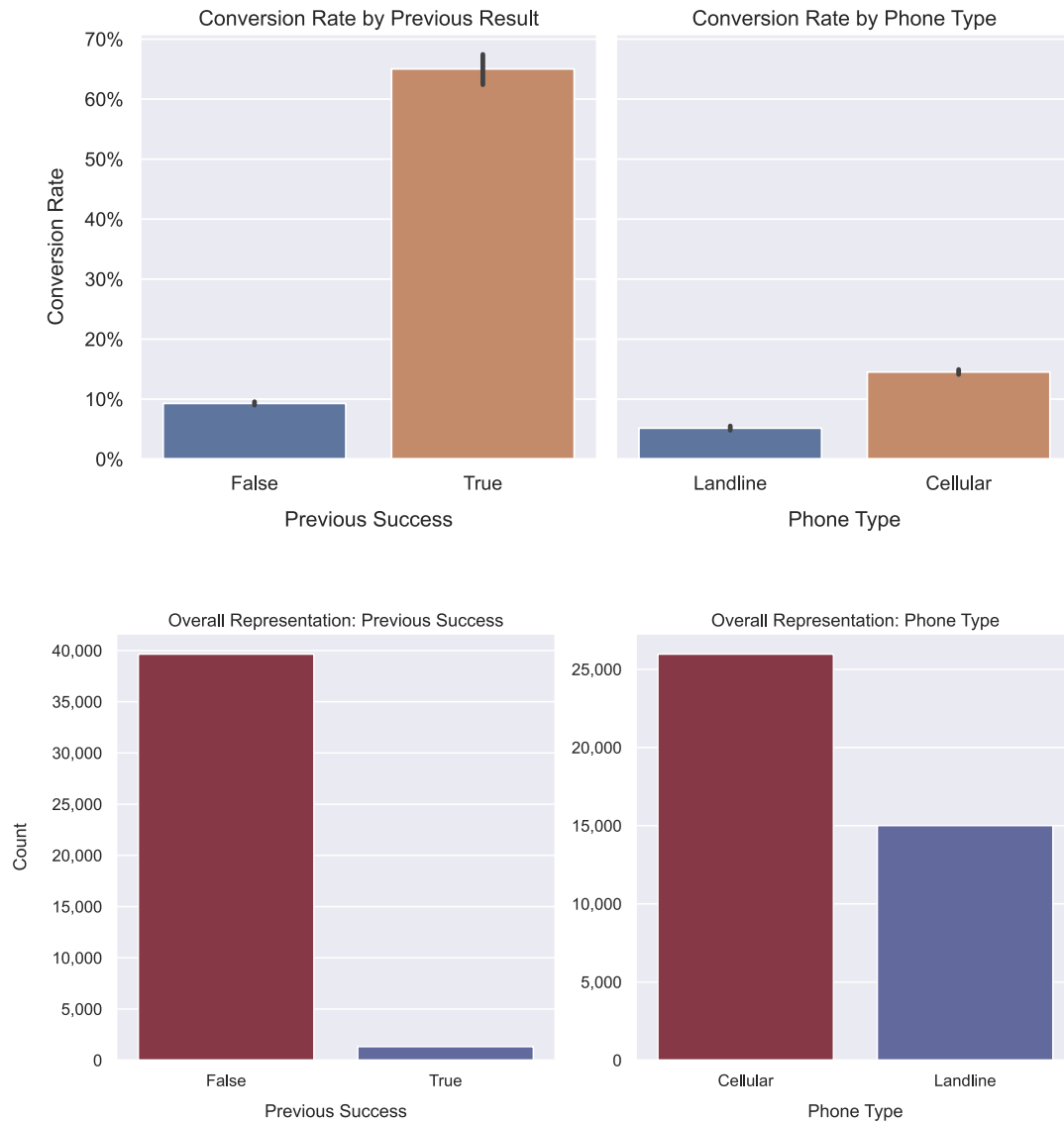
Previous Success is a Major Predictor

- Previous success is one of the biggest predictors of future success.
- Unfortunately, only about ~14% of the customers were previously contacted, and of those only ~3% made a deal.
- Cell phones are more popular than landlines, but other than that its unclear why they have a relationship with conversion.

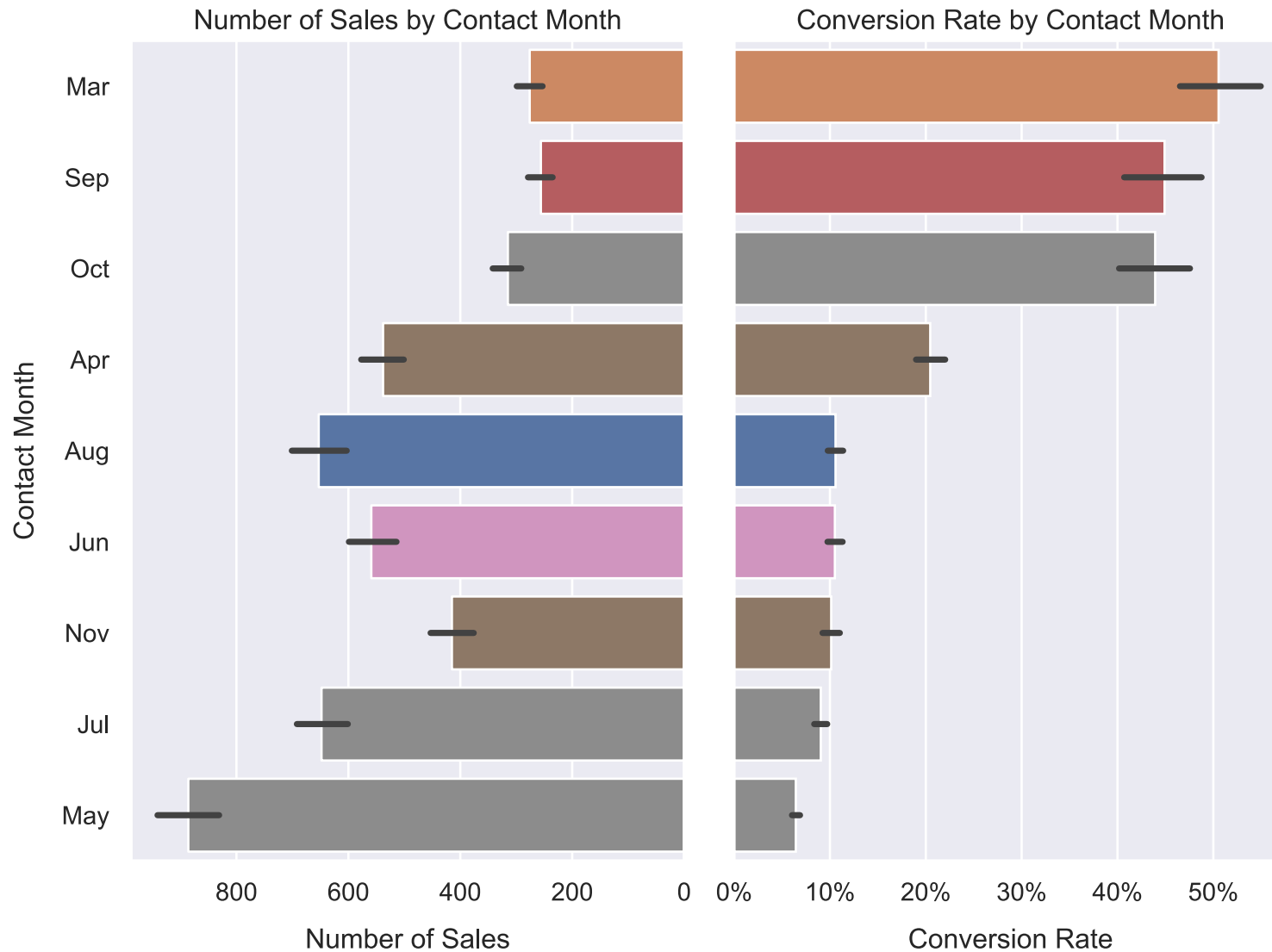
Recommendations:

Spend your energy and resources returning to previous customers who were receptive to marketing.

Focus on cell-users rather than landline-users.



Investment by Contact Month



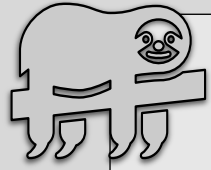
Relax in the Summer

- May has the highest total sales by the lowest average conversion rate
- It also has by far the most data points of any month, at nearly 14,000
- The summer months in general have the highest representation in the dataset

Recommendation:

Put less energy into summer marketing.

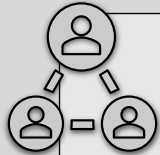
Recommendations Recap



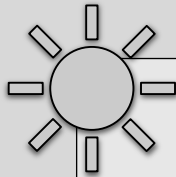
Pump resources into marketing when employment is **low**.



Relax your efforts when employment is **high**.



Return to previous customers who were receptive to marketing.



Put less energy into summer marketing.

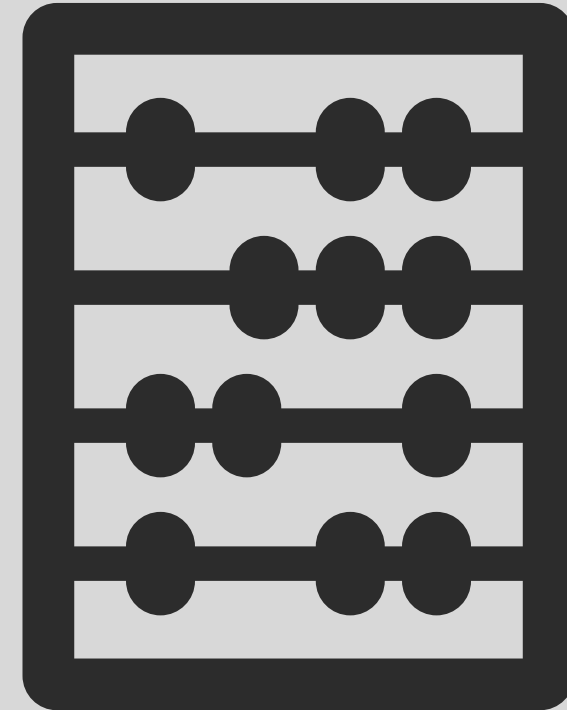
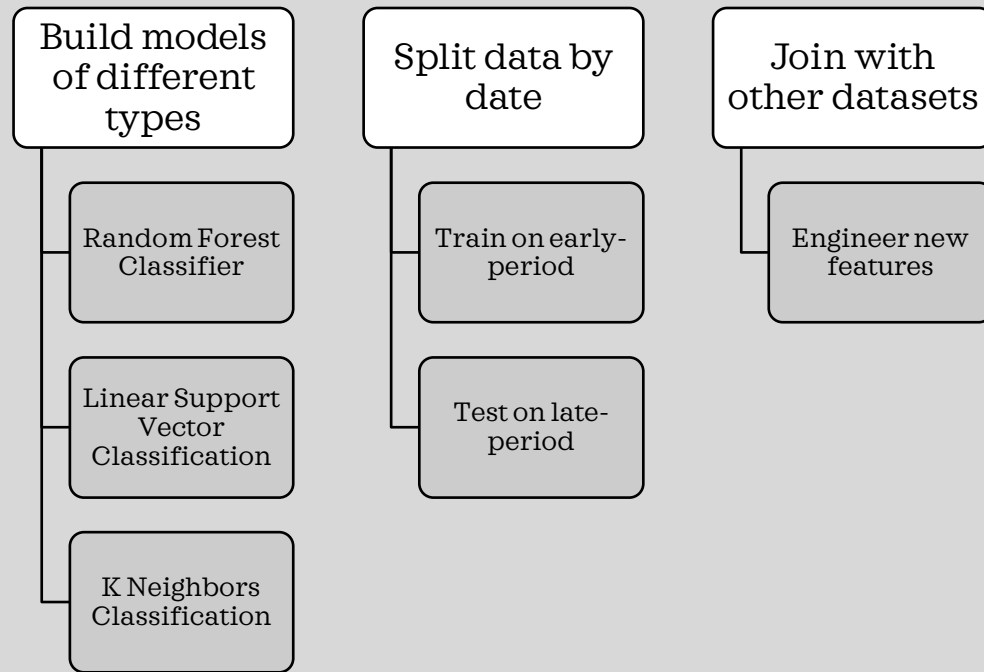


Focus on cell-users rather than landline-users.



**BANCO DE
PORTUGAL**
EUROSISTEMA

Future Work: More Models





Thank you!