



Thank you for accepting the rules.



\$60,000 • 884 teams

Santander Product Recommendation

Merger and Entry Deadline

Wed 26 Oct 2016

Wed 21 Dec 2016 (27 days to go)

Dashboard

Home

Data

Make a submission

Information

Description

Evaluation

Rules

Prizes

Timeline

Forum

Kernels

New Script

New Notebook

Leaderboard

My Team

My Submissions

Public Leaderboard

1. Tom Van de Wiele
2. Jack (Japan)
3. BreakfastPirate
4. Jamón de York de bellota
5. Alejo y Miro
6. ahaldenby
7. skyramp
8. raddar
9. alijs
10. Joshua Havelka

1,444 Kernels

Detailed Cleaning/Visualization (Python)
83 Votes / yesterday / Python

When Less is More.!
25 Votes / 3 days ago / Python

[Competition Details](#) » [Get the Data](#) » [Make a submission](#)

Data Files

| File Name | Available Formats |
|-----------------------|----------------------------------|
| sample_submission.csv | .zip (2.37 mb) |
| test_ver2.csv | .zip (12.39 mb) |
| train_ver2.csv | .zip (214.26 mb) |

In this competition, you are provided with 1.5 years of customers behavior data from Santander bank to predict what new products customers will purchase. The data starts at 2015-01-28 and has monthly records of products a customer has, such as "credit card", "savings account", etc. You will predict what **additional** products a customer will get in the last month, 2016-06-28, in addition to what they already have at 2016-05-28. These products are the columns named: ind_(xyz)_ult1, which are the columns #25 - #48 in the training data. You will predict what a customer will buy **in addition to what they already had at 2016-05-28**.

The test and train sets are split by time, and public and private leaderboard sets are split randomly.

Please note: This sample does not include any real Santander Spain customers, and thus it is not representative of Spain's customer base.

File descriptions

- **train.csv** - the training set
- **test.csv** - the test set
- **sample_submission.csv** - a sample submission file in the correct format

Data fields

| Column Name | Description |
|-------------|--|
| fecha_dato | The table is partitioned for this column |

Detailed Cleaning/Visualization
107 Votes / 12 days ago / R

Reduce Size of Dataset to 1 GB
11 Votes / 10 days ago / Python

Santander Products Visualisation
8 Votes / 6 days ago / Python

Santander quick first view
61 Votes / 13 days ago / R

Forum (119 topics)

Question about MAP@7
1 hour ago

LSTM?
2 hours ago

When Less is More.!
3 hours ago

[Errno 13] Permission denied
3 hours ago

When Less is More
9 hours ago

Messy predictors (age, antigüedad, ...)
11 hours ago

teams

players

entries

| Column Name | Description |
|-----------------------|---|
| ncodpers | Customer code |
| ind_empleado | Employee index: A active, B ex employed, F filial, N not employee, P pasive |
| pais_residencia | Customer's Country residence |
| sexo | Customer's sex |
| age | Age |
| fecha_alta | The date in which the customer became as the first holder of a contract in the bank |
| ind_nuevo | New customer Index. 1 if the customer registered in the last 6 months. |
| antigüedad | Customer seniority (in months) |
| indrel | 1 (First/Primary), 99 (Primary customer during the month but not at the end of the month) |
| ult_fec_cli_1t | Last date as primary customer (if he isn't at the end of the month) |
| indrel_1mes | Customer type at the beginning of the month ,1 (First/Primary customer), 2 (co-owner),P (Potential),3 (former primary), 4(former co-owner) |
| tiprel_1mes | Customer relation type at the beginning of the month, A (active), I (inactive), P (former customer),R (Potential) |
| indresi | Residence index (S (Yes) or N (No) if the residence country is the same than the bank country) |
| indext | Foreigner index (S (Yes) or N (No) if the customer's birth country is different than the bank country) |
| conyuemp | Spouse index. 1 if the customer is spouse of an employee |
| canal_entrada | channel used by the customer to join |
| indfall | Deceased index. N/S |
| tipodom | Addres type. 1, primary address |
| cod_prov | Province code (customer's address) |
| nomprov | Province name |
| ind_actividad_cliente | Activity index (1, active customer; 0, inactive customer) |
| renta | Gross income of the household |
| segmento | segmentation: 01 - VIP, 02 - Individuals 03 - college graduated |
| ind_ahor_fin_ult1 | Saving Account |
| ind_aval_fin_ult1 | Guarantees |
| ind_cco_fin_ult1 | Current Accounts |
| ind_cder_fin_ult1 | Derivada Account |
| ind_cno_fin_ult1 | Payroll Account |
| ind_ctju_fin_ult1 | Junior Account |
| ind_ctma_fin_ult1 | Más particular Account |
| ind_ctop_fin_ult1 | particular Account |
| ind_ctpp_fin_ult1 | particular Plus Account |
| ind_deco_fin_ult1 | Short-term deposits |
| ind_deme_fin_ult1 | Medium-term deposits |
| ind_dela_fin_ult1 | Long-term deposits |

| Column Name | Description |
|-------------------|--------------|
| ind_ecue_fin_ult1 | e-account |
| ind_fond_fin_ult1 | Funds |
| ind_hip_fin_ult1 | Mortgage |
| ind_plan_fin_ult1 | Pensions |
| ind_pres_fin_ult1 | Loans |
| ind_reca_fin_ult1 | Taxes |
| ind_tjcr_fin_ult1 | Credit Card |
| ind_valo_fin_ult1 | Securities |
| ind_viv_fin_ult1 | Home Account |
| ind_nomina_ult1 | Payroll |
| ind_nom_pens_ult1 | Pensions |
| ind_recibo_ult1 | Direct Debit |