# **A社データセットの概要**

## 1.データベースについて

* 電気通信事業会社A社の過去のデータが含まれたデータベースファイル「telecom.sqlite3」が配布されている
* カラム名とその説明が明らかになっている（「4-3.カラムの説明」参照）
* 以下の情報が格納された2種類のテーブルが含まれている
  1. 顧客情報（Client）
  2. 利用履歴（Record）
* データセットの特徴
  + 合計100000レコード×約100カラムのサイズの大きいテーブルデータ
    - EDAをしっかり行い、データについて理解する必要がある
    - コンペで扱ったデータと似たような性質を持つ
  + 最初にSQLを用いてデータを取得する
    - サンプルコードを配布しておりますので、どなたでもお気軽に取り組むことができます
* こんな方におすすめ
  + コンペで培ってきたEDAやモデリング力を試したい
  + SQLの復習をしたい

## 2.データベースの利用

1. SQLiteのインストールができている場合
   1. SQLiteを用いて、データベースに接続
   2. 各自、データベースを観察
   3. データ分析、事業提案に必要なデータをcsv形式で出力
   4. 各自必要なpythonモジュール（pandasなど）でデータ分析
2. SQLiteのインストールができていない場合
   1. pythonの標準ライブラリを用いて、Aと同様のことを実行する
3. 何から手をつければ良いか分からない場合
   1. 参考資料として配布されているsample codeを使用する
   2. 全データをcsv形式で出力
   3. 各自必要なpythonモジュール（pandasなど）でデータ分析

## 3.カラムの説明

　A社の担当者にデータについて問い合わせたところ、以下の説明が送られてきました。現段階では正式なコンサルティング契約は未締結であり、情報保護規程の厳しいA社からこれ以上の詳しい情報開示を受けることはできません。

　カテゴリカル変数の各種記号や略語の意味など不明な点は多々ありますが、不明な点があること自体も業務上の与条件として分析を進めてください。

| カラム名 | 内容の説明 |
| --- | --- |
| rev\_Mean | Mean monthly revenue (charge amount) |
| mou\_Mean | Mean number of monthly minutes of use |
| totmrc\_Mean | Mean total monthly recurring charge |
| da\_Mean | Mean number of directory assisted calls |
| ovrmou\_Mean | Mean overage minutes of use |
| ovrrev\_Mean | Mean overage revenue |
| vceovr\_Mean | Mean revenue of voice overage |
| datovr\_Mean | Mean revenue of data overage |
| roam\_Mean | Mean number of roaming calls |
| change\_mou | Percentage change in monthly minutes of use vs previous three month average |
| change\_rev | Percentage change in monthly revenue vs previous three month average |
| drop\_vce\_Mean | Mean number of dropped (failed) voice calls |
| drop\_dat\_Mean | Mean number of dropped (failed) data calls |
| blck\_vce\_Mean | Mean number of blocked (failed) voice calls |
| blck\_dat\_Mean | Mean number of blocked (failed) data calls |
| unan\_vce\_Mean | Mean number of unanswered voice calls |
| unan\_dat\_Mean | Mean number of unanswered data calls |
| plcd\_vce\_Mean | Mean number of attempted voice calls placed |
| plcd\_dat\_Mean | Mean number of attempted data calls placed |
| recv\_vce\_Mean | Mean number of received voice calls |
| recv\_sms\_Mean | NaN |
| comp\_vce\_Mean | Mean number of completed voice calls |
| comp\_dat\_Mean | Mean number of completed data calls |
| custcare\_Mean | Mean number of customer care calls |
| ccrndmou\_Mean | Mean rounded minutes of use of customer care calls |
| cc\_mou\_Mean | Mean unrounded minutes of use of customer care (see CUSTCARE\_MEAN) calls |
| inonemin\_Mean | Mean number of inbound calls less than one minute |
| threeway\_Mean | Mean number of three way calls |
| mou\_cvce\_Mean | Mean unrounded minutes of use of completed voice calls |
| mou\_cdat\_Mean | Mean unrounded minutes of use of completed data calls |
| mou\_rvce\_Mean | Mean unrounded minutes of use of received voice calls |
| owylis\_vce\_Mean | Mean number of outbound wireless to wireless voice calls |
| mouowylisv\_Mean | Mean unrounded minutes of use of outbound wireless to wireless voice calls |
| iwylis\_vce\_Mean | NaN |
| mouiwylisv\_Mean | Mean unrounded minutes of use of inbound wireless to wireless voice calls |
| peak\_vce\_Mean | Mean number of inbound and outbound peak voice calls |
| peak\_dat\_Mean | Mean number of peak data calls |
| mou\_peav\_Mean | Mean unrounded minutes of use of peak voice calls |
| mou\_pead\_Mean | Mean unrounded minutes of use of peak data calls |
| opk\_vce\_Mean | Mean number of off-peak voice calls |
| opk\_dat\_Mean | Mean number of off-peak data calls |
| mou\_opkv\_Mean | Mean unrounded minutes of use of off-peak voice calls |
| mou\_opkd\_Mean | Mean unrounded minutes of use of off-peak data calls |
| drop\_blk\_Mean | Mean number of dropped or blocked calls |
| attempt\_Mean | Mean number of attempted calls |
| complete\_Mean | Mean number of completed calls |
| callfwdv\_Mean | Mean number of call forwarding calls |
| callwait\_Mean | Mean number of call waiting calls |
| churn | Instance of churn between 31-60 days after observation date |
| months | Total number of months in service |
| uniqsubs | Number of unique subscribers in the household |
| actvsubs | Number of active subscribers in household |
| new\_cell | New cell phone user |
| crclscod | Credit class code |
| asl\_flag | Account spending limit |
| totcalls | Total number of calls over the life of the customer |
| totmou | Total minutes of use over the life of the customer |
| totrev | Total revenue |
| adjrev | Billing adjusted total revenue over the life of the customer |
| adjmou | Billing adjusted total minutes of use over the life of the customer |
| adjqty | Billing adjusted total number of calls over the life of the customer |
| avgrev | Average monthly revenue over the life of the customer |
| avgmou | Average monthly minutes of use over the life of the customer |
| avgqty | Average monthly number of calls over the life of the customer |
| avg3mou | Average monthly minutes of use over the previous three months |
| avg3qty | Average monthly number of calls over the previous three months |
| avg3rev | Average monthly revenue over the previous three months |
| avg6mou | Average monthly minutes of use over the previous six months |
| avg6qty | Average monthly number of calls over the previous six months |
| avg6rev | Average monthly revenue over the previous six months |
| prizm\_social\_one | Social group letter only |
| area | Geogrpahic area |
| dualband | Dualband |
| refurb\_new | Handset: refurbished or new |
| hnd\_price | Current handset price |
| phones | Number of handsets issued |
| models | Number of models issued |
| hnd\_webcap | Handset web capability |
| truck | Truck indicator |
| rv | RV indicator |
| ownrent | Home owner/renter status |
| lor | Length of residence |
| dwlltype | Dwelling Unit type |
| marital | Marital Status |
| adults | Number of adults in household |
| infobase | InfoBase match |
| income | Estimated income |
| numbcars | Known number of vehicles |
| HHstatin | Premier household status indicator |
| dwllsize | Dwelling size |
| forgntvl | Foreign travel dummy variable |
| ethnic | Ethnicity roll-up code |
| kid0\_2 | Child 0 - 2 years of age in household |
| kid3\_5 | Child 3 - 5 years of age in household |
| kid6\_10 | Child 6 - 10 years of age in household |
| kid11\_15 | Child 11 - 15 years of age in household |
| kid16\_17 | Child 16 - 17 years of age in household |
| creditcd | Credit card indicator |
| eqpdays | Number of days (age) of current equipment |
| Customer\_ID | NaN |