DATA DESCRIPTION: THE ITALIAN RESTAURANT CHAIN

This dataset includes extremely current data that allows us to study how to model retail data, how to incorporate the information on online and offline communication investments, and how to study a geographical distributed business. The company is interested in increasing store traffic and store sales. Not all stores operate during the entire period under analysis.

The Italian food chain data includes the following files:

1. **Performance data (Chain_Store_Performance.csv)**

   This file contains the data on the Net Euro Sales and Number of Receipts per store for Italy.
   The fields included are self-explanatory:

Field	Description
BUSINESS_DATE	day (short date format)
RESTAURANT_CODE	code for the restaurant (unique for each restaurant)
POSTCODE	Italian CAP or Postcode (in some cases we will have multiple restaurants per postcode)
NET_EURO_SALES	Total sales in euros per store and per day
RECEIPTS_COUNT	Number of receipts per store and per day (an approximate measure of traffic though not completely accurate)

2. **TV and Radio GRPs (Chain_GRPS_2015_2016.csv)**

   The chain has invested in multiple communication activities during the period under analysis. This file contains the total gross rating points (GRPs) that correspond to the investment for each day for the entire country (we do not have any local numbers available).

   The fields are also self-explanatory:

Field	Description
Date	day (short date format)
TV_GRPS	TV GRPs associated with chain's investment
Radio_GRPS	Radio GRPs associated with chain's investment

   GRPS are only available for the year of 2015 and 2016 (no data currently available for the first months of 2017). This is the simplest way to incorporate communication investments in a model.

3. Detailed chain investments (Campaign_Details.csv)

Although GRPs are a way of measuring the communication investments, they are not the only way. We can also use the amount spent (in euros) for each campaign or the number of campaigns active in a given day. We could even consider not only the offline communication investments but also the online ones.

This file contains the details of each the specific communication campaign.

Field	Description
CAMPAIGN_ID	The id number of the campaign
MEDIA	The media used (whether print, TV, internet and so on)
NETWORK	The Media Network used (only relevant for TV or similar)
CHANNEL_VEHICLE	Channel or media vehicle adopted (valid for all)
FORMAT	The format of the campaign (includes the seconds for TV ads, whether video or the type of display ads)
ISSUE_TIME	Either the issue (e.g., of the print magazine) or the time of the day (e.g., for a TV ad)
NET_COST	The cost in Euros of the campaign (some are zero or very low, see below for explanation)
START_DATE	Start date of the campaign
END_DATE	End date of the campaign

Important to consider the following: sometimes the cost of a campaign is zero or very low. This means that a specific communication campaign took place but because it was included in a special offer or, because of other arrangements, we cannot determine the exact cost of a campaign. When summing investments and using the euros spent in communication as a dependent variable, we do not consider these campaigns.

Maybe you want to think of ways of adding information to your model so that these zero cost campaigns can be reflected in the data. In the case of TV, zero cost campaigns are included in the GRPs.

4. Google Trends Data (GoogleTrends.csv)

Google trend scores obtained for the Italian market. The fields of this dataset include:

Field	Description
Week	The starting day of the week
GT_Chain1	The Google Trend score for Chain 1 (Focal Chain)
GT_Chain2	The Google Trend score for Chain 2 (Competitor 1)
GT_Chain3	The Google Trend score for Chain 3 (Competitor 2)
GT_Chain4	The Google Trend score for Chain 4 (Competitor 3)