MA-Project Data

Here are some data that you could consider for your final project.

|  |  |  |
| --- | --- | --- |
| **Title** | **Data Description** | **Data File** |
| Shopper Card Data | [Data Dictionary from dunnhumby](https://darden.instructure.com/courses/2146314/files/111708561/download?wrap=1) | [Carbo Loading](https://darden.instructure.com/courses/2146314/files/111708558/download?wrap=1)  [Complete Journey](https://darden.instructure.com/courses/2146314/files/111708560/download?wrap=1) |
| Online Retail Service Quality | Online retail prices, service quality ratings, merchant website traffic, and merchant descriptors from Bizrate.com | [Online Retail Price and Service Quality Rating Data](https://darden.instructure.com/courses/2146314/files/111709610/download?wrap=1) |
| Pharmaceutical Data | Monthly physician drug prescriptions, monthly sales calls allocated per physician, and survey responses on sales person and drug quality for a single therapeutic category. Data is available for over 48 months. | [Pharma Data](https://darden.instructure.com/courses/2146314/files/111709611/download?wrap=1) |
| Brand Marketing Mix Data | Annual unit sales, price, and advertising investments for all vodka brands in the U.S. for more than 15 years. | [Vodka Data](https://darden.instructure.com/courses/2146314/files/111709614/download?wrap=1) |
| Mobile Data | Mobile Coupon Distribution and Redemption Data from Cardagin. | [Cardagin Data](https://darden.instructure.com/courses/2146314/files/111709609/download?wrap=1) |
| Scanner Data | Store level, monthly brand (&sku) sales, price, promotion, feature/display for 1000+ brands in over 30 categories from 2000 until 2006. | [Marketing Science article: "The IRI Marketing Data Set" describing the data](https://darden.instructure.com/courses/2146314/files/111709613/download?wrap=1)  This is a very large data set that does not open in excel.  Contact me for help with accessing this data from UVA's cloud cluster.  You will need R or SAS to work with this data. |
| Emerging Market  Sales Data | Sales and marketing mix data by sku for all the brands in the beverage category across retail formats spanning mom and pop, small supermarket, and large supermarket in an emerging economy. | [Beverage Data Mom and Pop](https://darden.instructure.com/courses/2146314/files/111709607/download?wrap=1)  [Beverage Data Small Super Market](https://darden.instructure.com/courses/2146314/files/111709608/download?wrap=1)  [Beverage Data Large Super Market](https://darden.instructure.com/courses/2146314/files/111709606/download?wrap=1) |
| Movie Preference Data | Movie preference ratings for 2000+ consumers. | [Movie Preference Data ReadMe](https://darden.instructure.com/courses/2146314/files/111712049/download?wrap=1)  [Genres](https://darden.instructure.com/courses/2146314/files/111711962/download?wrap=1)  [Data Size Summary](https://darden.instructure.com/courses/2146314/files/111711935/download?wrap=1)  [Movie Information](https://darden.instructure.com/courses/2146314/files/112139018/download?wrap=1)  [Occupation Information](https://darden.instructure.com/courses/2146314/files/111712072/download?wrap=1)  [User Information](https://darden.instructure.com/courses/2146314/files/111712165/download?wrap=1)  [Masterdata](https://darden.instructure.com/courses/2146314/files/111711994/download?wrap=1)  [Sample Train Data](https://darden.instructure.com/courses/2146314/files/111712115/download?wrap=1)  [Sample Test Data](https://darden.instructure.com/courses/2146314/files/111712085/download?wrap=1) |
| Mobile Keyword Data | Data that links keyword searches, and Xbox game   purchases for best buy.com.  Information on the data files is provided in the following link: https://www.kaggle.com/c/acm-sf-chapter-hackathon-big/data | [Best Buy Training Data](https://darden.instructure.com/courses/2146314/files/111713521/download?wrap=1)  [Best Buy Test Data](https://darden.instructure.com/courses/2146314/files/111713519/download?wrap=1)  [Best Buy Sku Description](https://darden.instructure.com/courses/2146314/files/112138787/download?wrap=1) |
| Rotten Tomatoes Dataset | **Classify the sentiment of sentences from the Rotten Tomatoes dataset.**  The Rotten Tomatoes movie review dataset is a corpus of movie reviews used for sentiment analysis.  This data presents a chance to benchmark your sentiment-analysis ideas on the Rotten Tomatoes dataset.  You are asked to label phrases on a scale of five values: negative, somewhat negative, neutral, somewhat positive, or positive. Obstacles like sentence negation, sarcasm, terseness, language ambiguity, and many others make this task very challenging. | [Train Data](https://darden.instructure.com/courses/2146314/files/111712115/download?wrap=1)  [Test Data](https://darden.instructure.com/courses/2146314/files/111712085/download?wrap=1) |
| Multi-channel Marketing | Data is from a multichannel company with sales of several hundred million dollars per year. This nationally known company has a network of retail stores, a well-established traditional catalog channel, and a website.   The majority of its sales are food products purchased as gifts during the Christmas season. This data set includes over 100,000 customer records and over 3.3 million marketing contact records. | [Demo Codes](https://darden.instructure.com/courses/2146314/files/111713523/download?wrap=1)  [Data Description](https://darden.instructure.com/courses/1857279/files/86968115/download?wrap=1)  [Data](https://darden.instructure.com/courses/2146314/files/112138967/download?wrap=1)    This is a very large data set that does not open in excel.  Contact me for help with accessing this data from UVA's cloud cluster.  You will need R or SAS to work with this data. |