1. Bigmart Sales Data Set

 Retail is another industry which extensively uses analytics to optimize business processes. Tasks like product placement, inventory management, customized offers, product bundling etc are being smartly handled using data science techniques. As the name suggests, this data comprises of transaction record of a sales store. This is a regression problem.

**Problem:** Predict the sales.

2. Loan Prediction Data Set

Among all industries, insurance domain has the largest use of analytics & data science methods. This data set would provide you enough taste of working on data sets from insurance companies, what challenges are faced, what strategies are used, which variables influence the outcome etc. This is a classification problem.

**Problem:** Predict if a loan will get approved or not.

3. Black Friday Data Set

This data set comprises of sales transactions captured at a retail store. It’s a classic data set to explore your feature engineering skills and day to day understanding from your shopping experience. It’s a regression problem.

**Problem:** Predict purchase amount.

4. Yelp Data Set

This data set is a part of round 8 of The Yelp Dataset Challenge. It comprises of nearly 200,000 images, provided in 3 json files of ~2GB. These images provide information about local businesses in 10 cities across 4 countries. We are required to find insights from data using cultural trends, seasonal trends, infer categories, text mining, social graph mining etc.

**Problem:** Find insights from images