Food Sales Predictions Project

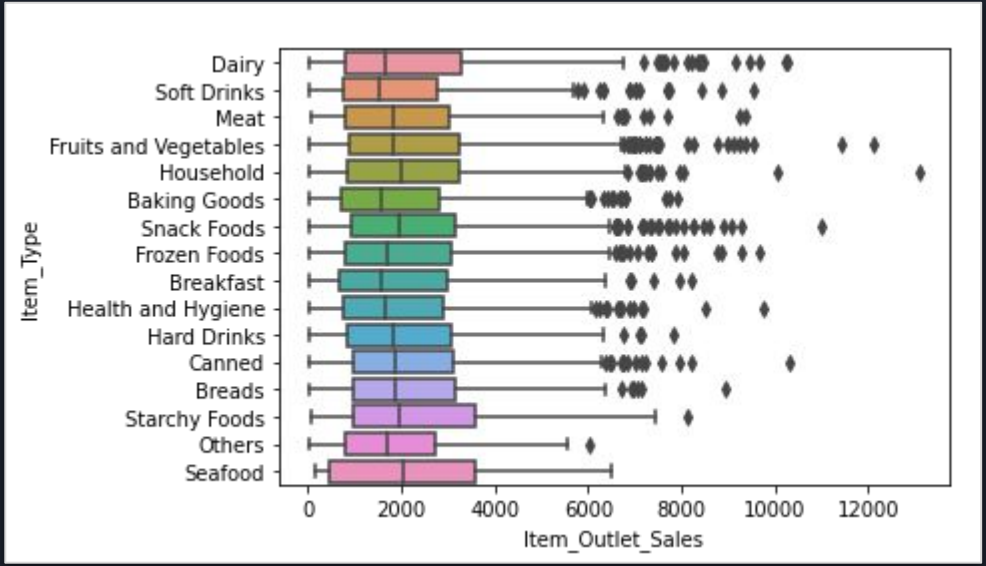
Madeline Ye - 8.30_Brenda_cohort

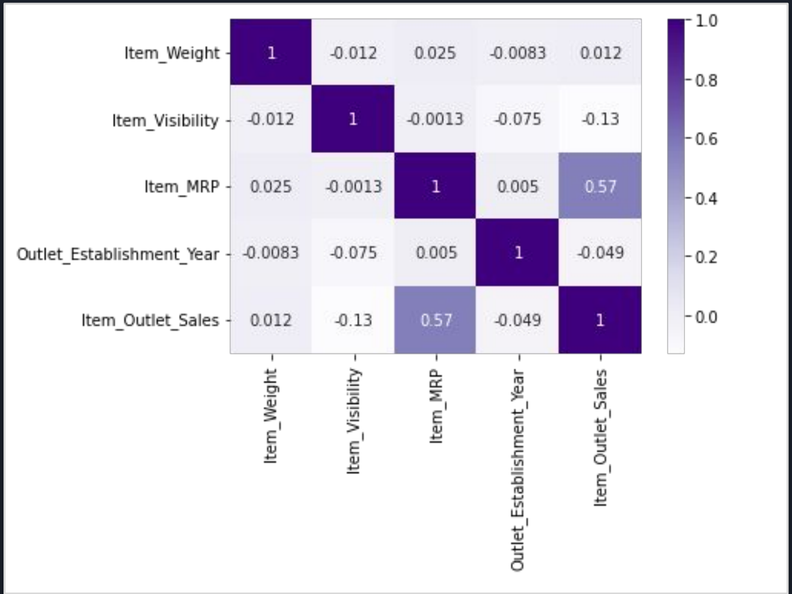
What, why, and how?

- Grocery and Supermarket sales data with columns of different features
- Explore and clean the data to make sure it's correct and ready to input into a Machine Learning Model
- Predict Sales after training and testing these said models to decide which performs best on this data set
    - Linear Regression
    - Regression Tree
    - Bagged Trees
    - Random Forest

	Item_Identifier	Item_Weight	Item_Fat_Content	Item_Visibility	Item_Type	Item_MRP	Outlet_Identifier	Outlet_Establishment_Year	Outlet_Size	Outlet_Location_Type	Outlet_Type	Item_Outlet_Sales
0	FDA15	9.30	Low Fat	0.016047	Dairy	249.8092	OUT049	1999	Medium	Tier 1	Supermarket Type1	3735.1380
1	DRC01	5.92	Regular	0.019278	Soft Drinks	48.2692	OUT018	2009	Medium	Tier 3	Supermarket Type2	443.4228
2	FDN15	17.50	Low Fat	0.016760	Meat	141.6180	OUT049	1999	Medium	Tier 1	Supermarket Type1	2097.2700
3	FDX07	19.20	Regular	0.000000	Fruits and Vegetables	182.0950	OUT010	1998	missing	Tier 3	Grocery Store	732.3800
4	NCD19	8.93	Low Fat	0.000000	Household	53.8614	OUT013	1987	High	Tier 3	Supermarket Type1	994.7052

Visualizing the Data

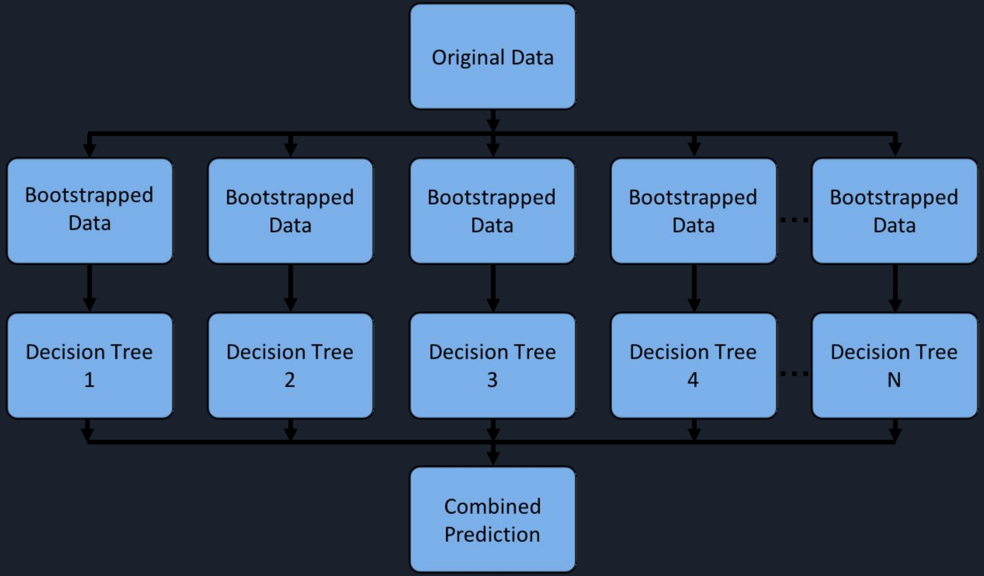


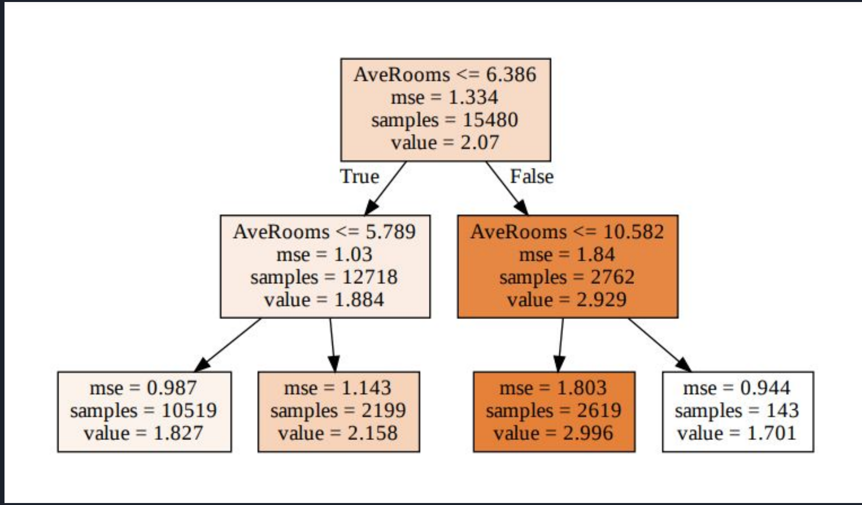


- Exploring the sales distribution of each item type
- Applying filters and running the same kind of chart (i.e. Different sales for different types of markets?)
- Any correlation between quantitative values?

Different Machine Learning Models

- Linear Regression: y = mx+b
- Decision/Regression Tree
- Bagged Trees
  - Bootstrapping + Aggregating
- Random Forests





Results

- Which model performed the best on our Food Sales data?
  - Decision Tree and Random Forests came in pretty close
- Measured based off of R2 value and RMSE value

- With the Random Forests Model:
  - We can conclude that 60% of the variation in the predictions can be accounted for by the features we selected
  - RMSE tells us that the error of our predictions will likely fall between $670, plus or minus for Item sales