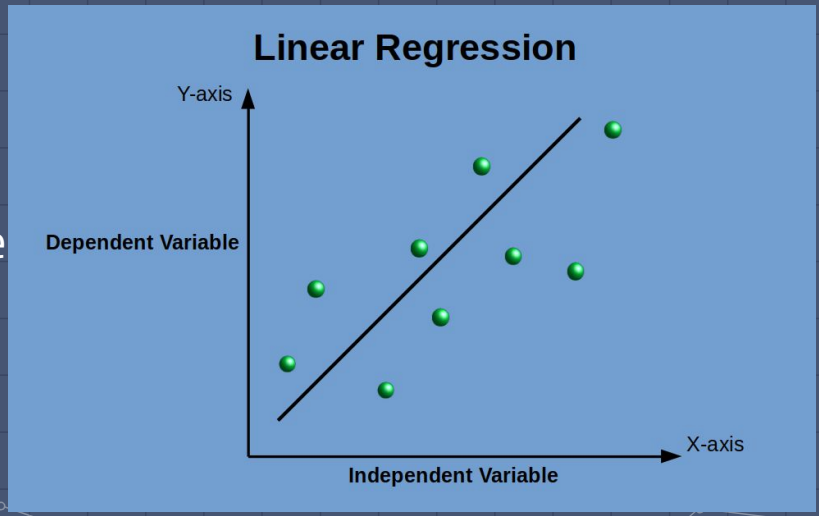
Applications of Linear Regression

What is Linear Regression?

- Very basic and commonly used machine learning model
- Used to predict the values of a variables give the values of other variables
- Line of best fit and Ordinary least squares
  - Provides the most accurate model



Reference: https://ai.plainenglish.io/linear-regression-clearly-explained-part-1-96dc1e9e70f7

Why do we use Linear Regression?

1. Linear Regression is very versatile and is widely applicational
2. Linear Regression is easily understandable and explained
3. Linear Regression is very simple to implement but still very powerful

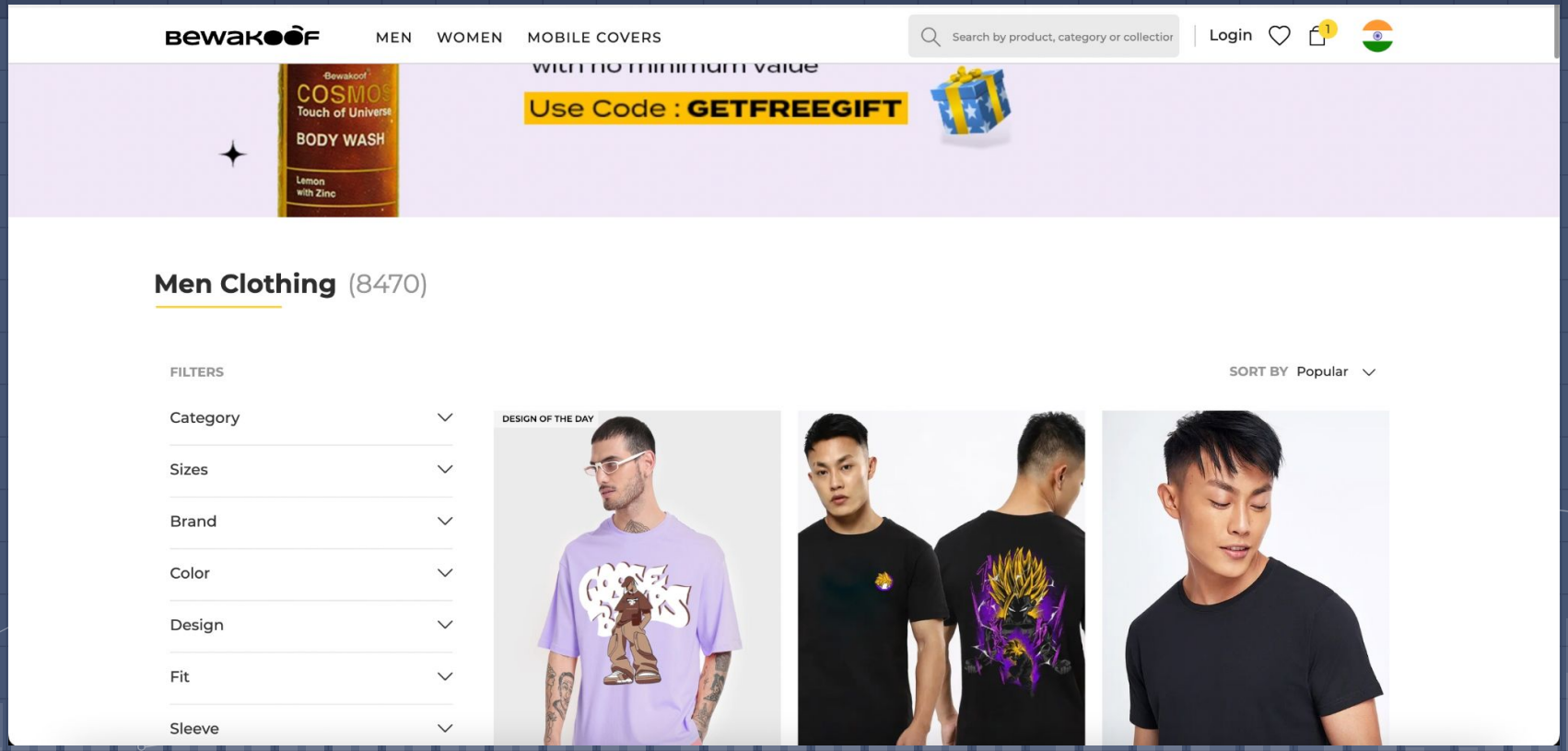
Examples of Real World Regression Models

- How much money should be allocated for gas
  - Predictors: MPG, miles travelling
- Predicting average grades of students in a class
- Predicting sales of a product

Reference: https://biblicalpreaching.net/2012/11/12/application-preaching/

Our Real World Applications

BEWAKOOF is an Online Shopping site for Men and Women Clothing.



Sales prediction

- Sales prediction is the process of predicting the amount of product or services a sales unit (which can be an individual salesperson, a sales team, or a company) will sell in the next week, month, quarter, or year.
- Predicting future sales for a company is one of the most important aspects of strategic planning.
- Applications
  - Inventory management
  - Product ranking algorithm

Key terms

- **Impressions**
  - A **Product Impression** is defined as viewing one (or usually multiple) product offers on a given page (usually called as **Product Listing Page or PLP**).
- **Clicks**
  - A **Product click** is defined as clicking on a product and going to the **Product Detail Page or PDP**.
- **Add to carts**
  - An **Add to cart** is defined as adding a product to cart/shopping bag from a **PDP**.
- **Sales**
  - Total units sold of a particular product.

Data

X

Y

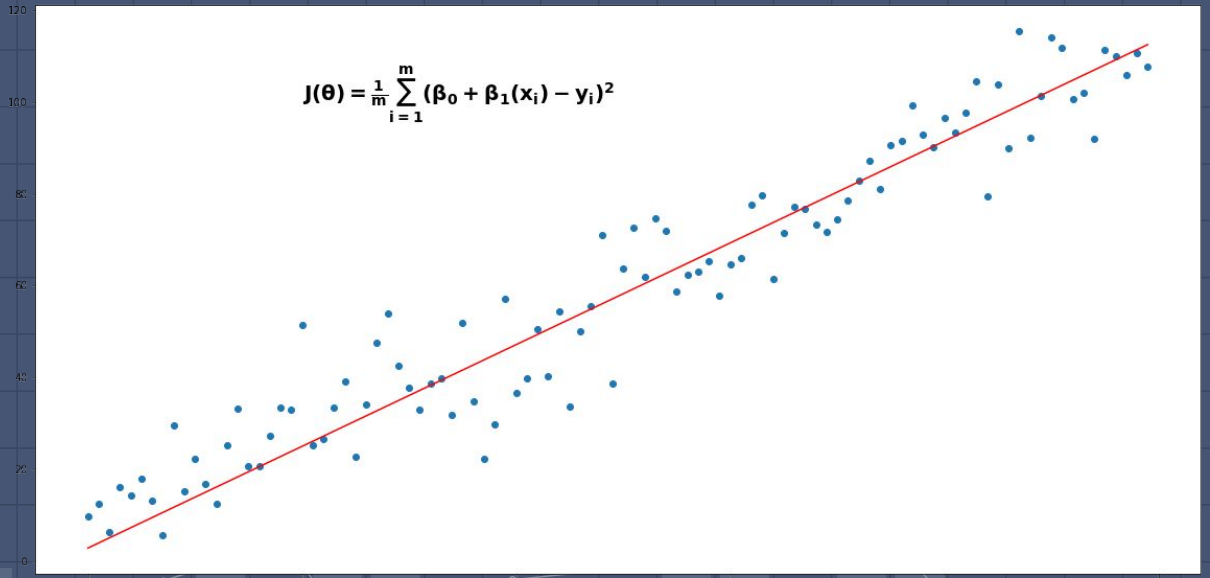
Product ID	Trailing 1 day imprs	Trailing 3 days imprs	Trailing 7 days imprs	Trailing 14 days imprs	Trailing 1 days clicks	....	Trailing 1 days a2cs			Leading 7 days sales
74441	12500	37500	87500	175000	250		5			1
74442	8900	26700	62300	124600	178	....	4			1
74443	10200	30600	71400	142800	204	....	4			1
74444	9250	27750	64750	129500	185	....	4			1
74445	15000	45000	105000	210000	300	....	6			2
						....				

Product ID	Gender	Category	Color	Price
74441	Men	T-Shirt	Black	299
74442	Women	Top	Red	349
74443	Women	Shorts	Black	449
74444	Men	T-Shirt	Blue	349
74445	Men	Shirt	White	649

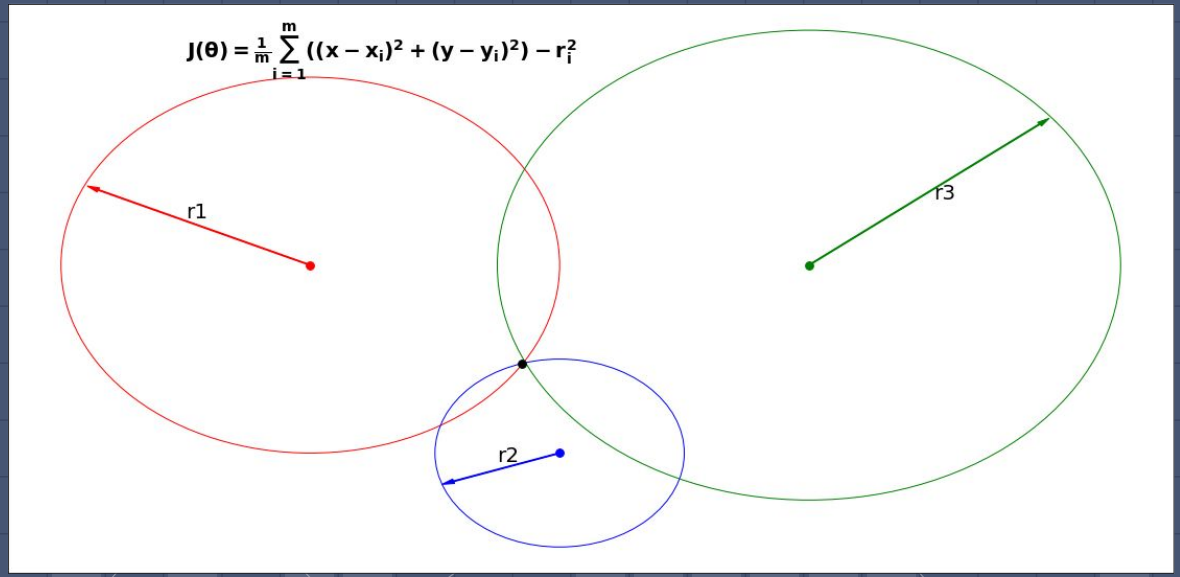
Conclusion

- Proved to be a very good baseline model
- There was a lot of scope of improvement not limited to including more predictors, checking for multicollinearity/ influential observations.
- Other algorithms used for prediction were KNN regressor, Random Forest regressor, XGBoost.

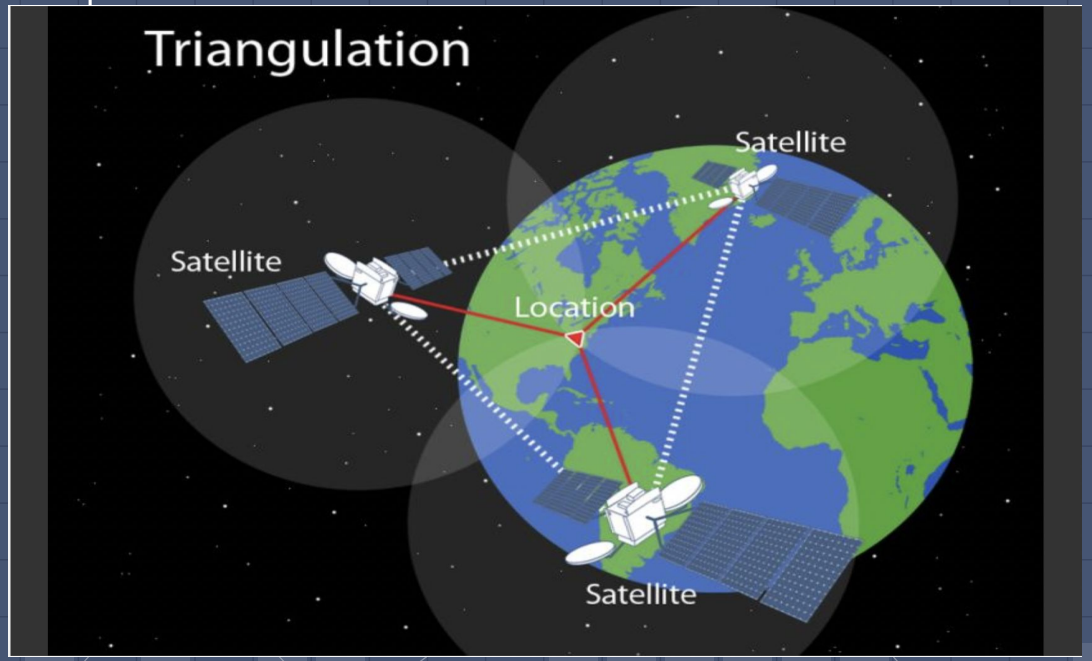
Least Squares to estimate best fit line



Least Squares to estimate best fit 2D point



Least Squares to estimate best fit 3D point



https://github.com/google/gps-measurement-tools/blob/master/opensource/WlsPvt.m

Big Picture

- Linear Regression is very popular regression method
  - Used for prediction
- Versatile and has wide applicability
- Less of a black box and easier to communicate
- Easy to implement

Thank You!