Project Proposal

Question/Need:

- What is the question behind your analysis or model and what practical impact will your work have?

Computer monitor is a major productivity tool for people whose work requires long time sitting in front of the computer. As a data practitioner or prospective data practitioner, you probably have or are searching for a right monitor for you. Best Buy as one of the top electronics retailer, offers multitudinous monitors to customers.

While as customers, we might feel overwhelmed when facing with so many features and so many options of monitors. One might wonders what feature(s) of monitors is(are) the main driving factor(s) of determining the prices of monitors? Thus, my analysis will focus on exploring what features of monitors affect the prices and, in turn, try to predict the price of a monitor given its features.

- Who is your client and how will that client benefits from exploring this question or building this model/system?

People who wants to buy monitors at Best Buy. My analysis on monitor prices will help people to learn about the value of each monitor features. And people might be able to avoid paying more than they planned or paying for some features that they don't care about.

Data Description:

- What dataset(s) do you plan to use, and how will you obtain the data? Please include a link! (The link can be to the dataset you're downloading, the site you're scraping, etc.)

I plan to scrap computer monitor listings at <u>Bestbuy.com</u> and create a dataset accordingly. I will scrap it twice, once a week for two weeks. Thus, I will have data of monitor listings from two different times that possibly shows the price changes over time. (The better case is to have more different times. Given this is

my first ever web scraping project and it is due in around two weeks, having data from two different timestamps seem to be the best practice at the point.)

- What is an individual sample/unit of analysis in this project? In other words, what does one row or observation of the data represent?

Basic information:

`Price` : price of computer monitor - **numerical**

`Model number` :model number of monitor that uniquely identifies a monitor - **categorical**

`Brand` : brand of monitor - **categorical**

`Rating` : Customer Reviews of the monitor - **numerical**

Monitor features:

Display:

`Refresh Rate` - numerical

`Maximum Resolution` - numerical

`Native Resolution` - numerical

`Contrast Ratio` - categorical

`Response Time` - numerical

`Display Type` - categorical

`Panel Type` - categorical

`Synchronization Technology` - categorical

`Aspect Ratio` - categorical

`Curved Screen` - YES/NO - categorical

`Brightness` - numerical

`Screen Size` - numerical

`High Dynamic Range (HDR)` - YES/NO - categorical

`Horizontal Frequency` - numerical

`Horizontal Viewing Angle` - numerical

`Vertical Viewing Angle` - numerical

`Smart Display` - YES/NO - categorical?";.l,kmjbhv cu

`Touch Screen` - YES/NO - categorical

`On-Screen Display Languages` - categorical

`Quantum Dot Technology` - YES/NO - categorical

Connectivity:

`Number of DVI Inputs` - numerical

`Number of VGA Ports` - numerical

`Number of HDMI Inputs (Total)` - numerical

`Number of HDMI Outputs (Total)` - numerical

`Number of DisplayPort Inputs (Total)` - numerical

`Number of Thunderbolt Ports (Total)` - numerical

`Number of USB Ports (Total)` - numerical

`Headphone Jack` - YES/NO - categorical

Compatibility:

`Voice Assistant Built-in` - YES/NO - categorical

`Wall Mountable` - YES/NO - categorical

Other:

`Built-In Webcam` - YES/NO - categorical

`Tiltable` - YES/NO - categorical

`Adjustable Stand Heigh` - YES/NO - categorical

`Integrated Speaker(s)` - YES/NO - categorical

`Anti-Glare` - YES/NO - categorical

`Weight` - numerical

`Power Consumption` - numerical

`Warranty` - numerical

- What characteristics/features do you expect to work with? In other words, what are your columns of interest?

My columns of interest would be all the above columns at the moment. I don't know what features are worth more focus so I will work on them to see and further narrow down to the features with more importance.

- If modeling, what will you predict as your target?

My target is the price of monitor.

Tools:

- How do you intend to meet the tools requirement of the project?

I will use Python for this project, specifically mainly with following libraries:

Web scraping: `BeautifulSoup` `selenium` `requests`

Data cleaning/analysis/modeling: `Pandas` `numpy` `sklearn`

Data Visualization: `matplotlib` `seaborn`

- Are you planning in advance to need or use additional tools beyond those required?

No

MVP Goal:

- What would a MVP Example look like for this project?

Graph shows relationship between Predicted price(best fitting line) and Actual prices from the baseline linear regression model.