
AiCore Project 1

Alex Henderson

Idea

A laptop's price derives from several factors:

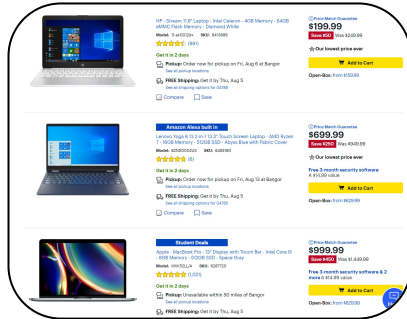
Continuous:

- Screen size
- Weight
- RAM/Internal Storage

Categorical:

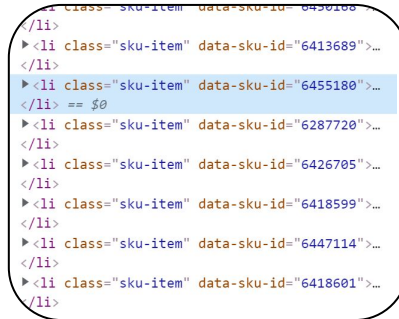
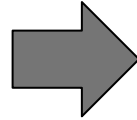
- Processor type

Scraping Flow



BestBuy.com

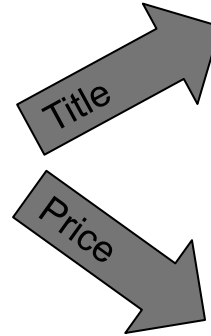
(Data also taken from other retail stored as such as Target.com)



Product cards stored in

Lenovo Yoga 6 13 2-in-1 13.3" Touch Screen Laptop - AMD Ryzen 7 - 16GB Memory - 512GB SSD - Abyss Blue with Fabric Cover

▼ <div class="sku-title">



```
span_elements = i.find_all("span")
```

```
if span[0] == '$':
```

Data Storage

- Data stored in a simple local Postgres database
 - Records are very structured, making other options such as a data lake/warehouse unnecessary
- Makes querying the data very quick & easy:

```
ai-core=# select *
ai-core=# from   laptop_data
ai-core=# where  price > 1000
ai-core=# order by price desc
ai-core=# limit  10;
```

title							price	screen_size	processor
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 64GB Memory	- 8TB SSD	- Space Gray		5899	16	i9
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 64GB Memory	- 8TB SSD	- Silver		5899	16	i9
Apple - MacBook Pro 16"	Laptop	- Intel Core i9	- 64GB Memory	- 8TB SSD	- Space Gray		5899	16	i9
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i7	- 64GB Memory	- 8TB SSD	- Space Gray		5799	16	i7
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 32GB Memory	- 8TB SSD	- Silver		5699	16	i9
Apple - MacBook Pro 16"	Laptop	- Intel Core i7	- 64GB Memory	- 8TB SSD	- Space Gray		5699	16	i7
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 32GB Memory	- 8TB SSD	- Space Gray		5699	16	i9
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 32GB Memory	- 8TB SSD	- Space Gray		5599	16	i9
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i7	- 64GB Memory	- 8TB SSD	- Silver		5599	16	i7
Apple - MacBook Pro 16"	Display with Touch Bar	- Intel Core i9	- 32GB Memory	- 8TB SSD	- Silver		5599	16	i9

(rows)

Data Cleaning/Pre-processing

- The raw data extracted from the web scraper comes in two strings, one containing the title of the laptop product, and one containing the price data.
- Can simply parse the text contained in these two strings to extract the specifications of the product:
 - Is the first character of the string '\$' - If so this is the price span.
 - Replace any commas or decimal points with spaces, and convert to floating points
 - If we have the title of the product, search for the different possible processors in the title string to identify which processor the product has

Next steps

- For the building of ML models etc..., a straightforward route to take would be to use laptop specs as the X variables, and price as the Y (target) variable.
 - Here the laptop specs can include a mixture of categorical and continuous variables (screen-size, processor type/brand, RAM/storage)



Other Notes

- I had separate project which scrapes tweets about bitcoin as well as the associated financial metrics, although I had problems with my Selenium Bot on the twitter site
- I applied for Twitter API access, and decided to abandon this direction for the time being