# Final Project Data Checkpoint

# Munjot Singh

Project Code:

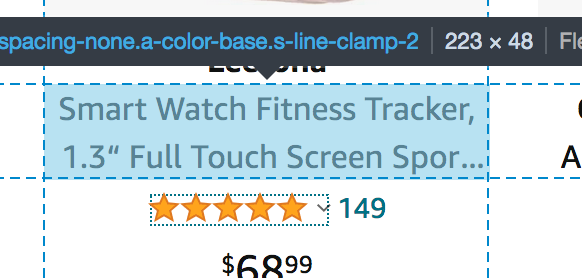
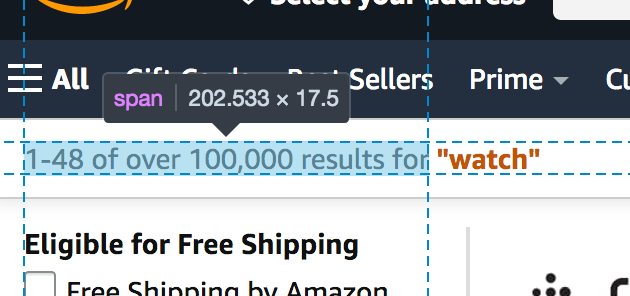
* <https://github.com/munjotks/FinalProject-Munjotks.git>

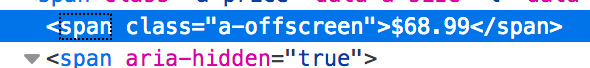
Data Sources:

* User will be searching for an Amazon product through the user interface. In order to retrieve the information from amazon, I will be scraping Amazon using cache 🡪 <https://www.amazon.com/>
* The URL for what page to scrape will be generated from <https://www.amazon.com/s?k=(SEARCHTERM)&ref=nb_sb_noss_1> and that specific page will be scraped.



* The data I will be collecting from the specific pages will be
  + Search Term | Product Name | # of star out of 5 | Product Price | # of Reviews | # of results





* Caching will be used every time the same search term is used in the user interface (Same URL)



Database:

* I will be creating a database from the information collected from scraping the search term pages. My two tables will consist of the following fields.
  + Product Table
    - Search Term Category | Product Name | # of Stars | # of Reviews
  + Category (search term) Table
    - Search Term Category | # of Results

EXAMPLE:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Search Term Category | Product Name | Product Price | # of Stars | # of Reviews |
| Camera | Fujifilm Instax Mini 11 Instant Camera - Lilac Purple | 69.00 | 4.8 | 2,528 |
| Camera | Digital Camera, Lecran FHD 1080P 36.0 Mega Pixels Vlogging Camera with 16X Digital Zoom, LCD Screen, Compact Portable | 48.99 | 4.3 | 83 |
| Camera | All-new Blink Outdoor – wireless, weather-resistant HD security camera with two-year battery life and motion detection | 59.99 | 4.3 | 4,619 |
| Camera | Digital Camera, Lecran FHD 1080P 36.0 Mega Pixels Vlogging Camera | 82.98 | 4.2 | 83 |

EXAMPLE:

|  |  |
| --- | --- |
| Search Term Category | # of Results |
| Camera | 10,000+ |
| Pen | 3,000 |

Interaction and Presentation Plans

User will be asked:

* What would you like to search on Amazon?
* User Response: Camera
* How would you like the results displayed?
* User Response: top 10 by stars
* Displays top 10 results
* User can request a scatterplot;
  + Product price vs. # of stars
  + Product price vs. # of reviews