

# Amazon Cell Phone Reviews

---

KuanHsun Lu  
YiCheng Chen  
FengJie Chen



# Background

---

- > When a customer wants to buy a new cell phone, he or she may wonder which product has good reputation and popularity. Our primary goal is to provide users with a platform where they can refer to. The platform is a web page, which includes the following features:
  - Visualization of sales volume for cell phones of all brands
  - Comparisons of ratings for different brands.
  - Display of most helpful reviews for a selected type.



# Data source

---

- > Our data comes from Kaggle open dataset, which are collected from Amazon. It consists of two data sets:
  - items.csv: contains item-level information, main attributes include product id(key), brand, title, rating, totalReviews.
  - reviews.csv: gathers user-level reviews for cell phones they bought, main attributes include product id(key), rating, review content, helpful votes.
- > Limitations
  - It cannot reflect the most recent market trend.
  - Not sufficient features.

# Brief introduction to Dataset

## > Item.csv

- Brand, title, url, image, rating reviewURL, total review, prices

A brand	A title	url	image	# rating	reviewUrl
Product Brand	Product Title	Product URL	Product Image URL	Product Avg. Rating	Product Review P
<b>Samsung</b> 50% <b>Apple</b> 13% Other (8) 37%	785 unique values	792 unique values	719 unique values		792 unique val
Nokia	Dual-Band / Tri-Mode Sprint PCS Phone w/ Voice Activated Dialing & Bright White Backlit Screen	https://www.amazon.com/Dual-Band-Tri-Mode-Activated-Dialing-Backlit/dp/B0000SX2UC	https://m.media-amazon.com/images/I/2143EBQ210L._AC_UY218_SEARCH213888_FMwebp_QL75_.jpg	3	https://www.amazon.com/product-reviews/B0000S
Motorola	Motorola I265 phone	https://www.amazon.com/Motorola-i265-I265-phone/dp/B0009N5L7K	https://m.media-amazon.com/images/I/419WBAVDARL._AC_UY218_SEARCH213888_FMwebp_QL75_.jpg	2.9	https://www.amazon.com/product-reviews/B0009N
Motorola	MOTOROLA C168i AT&T CINGULAR PREPAID GOPHONE CELL PHONE	https://www.amazon.com/MOTOROLA-C168i-CINGULAR-PREPAID-GOPHONE/dp/B000SKTZ0S	https://m.media-amazon.com/images/I/71b+q3ydkIS._AC_UY218_SEARCH213888_FMwebp_QL75_.jpg	2.6	https://www.amazon.com/product-reviews/B000SK
Nokia	Nokia 6500 Slide Black/silver Unlocked Cell Phone	https://www.amazon.com/Nokia-6500-Slide-silver-Unlocked/dp/B00198M12M	https://m.media-amazon.com/images/I/41ss4HpLkLL._AC_UY218_SEARCH213888_FMwebp_QL75_.jpg	2.4	https://www.amazon.com/product-reviews/B00198



# Brief introduction to Dataset

## > Review.csv

– Name, rating, date, review title, body

A name ▾	# rating ▾	📅 date ▾	✓ verified ▾	A title ▾	A body
Reviewer Name	Reviewer Rating (scale 1 to 5)	Review Date	Valid Customer	Review Title	Review Content
<b>Amazon Customer</b> 11% <b>Kindle Customer</b> 1% <b>Other (55974)</b> 89%				<b>Five Stars</b> 10% <b>One Star</b> 2% <b>Other (49135)</b> 88%	<b>72632</b> <b>unique val</b>
Janet	3	October 11, 2005	false	Def not best, but not worst	I had the Sams A600 for awhile which is absolute doo doo. You could read my review and detect my at the stupid. It finally died me so I used the Nokia phone I in a garage ..
Luke Wyatt	1	January 7, 2004	false	Text Messaging Doesn't Work	Due to a software issue between and Sprint this phone's text messaging capabilities could not work with Sprint system and worked until a software patch comes out "some time in



## Use cases - Visualization of sales amount and ratings for all brands

### > Input

- Users can select a brand of cell phone through a dropdown box

### > Output

- Stacked histogram of ratings for all brands
- Pie chart of sales amount and rating for a specific brand



## Use cases - Retrieve helpful reviews from our data

---

- > Input
  - Users can select a brand through a dropdown box
- > Output
  - Show all types of cell phones under the brand in a dropdown box
- > Input
  - After users select a brand, they can further select a specific type
- > Output
  - Retrieve and display the review with highest votes under this type.

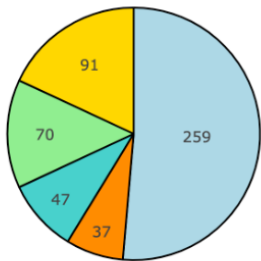


# Demo

If you are interested in ASUS, then choose ASUS.

ASUS

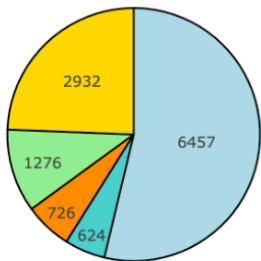
Pie chart of ASUS's ratings



If you are interested in Apple, then choose Apple.

Apple

Pie chart of Apple's ratings





# Design – software components

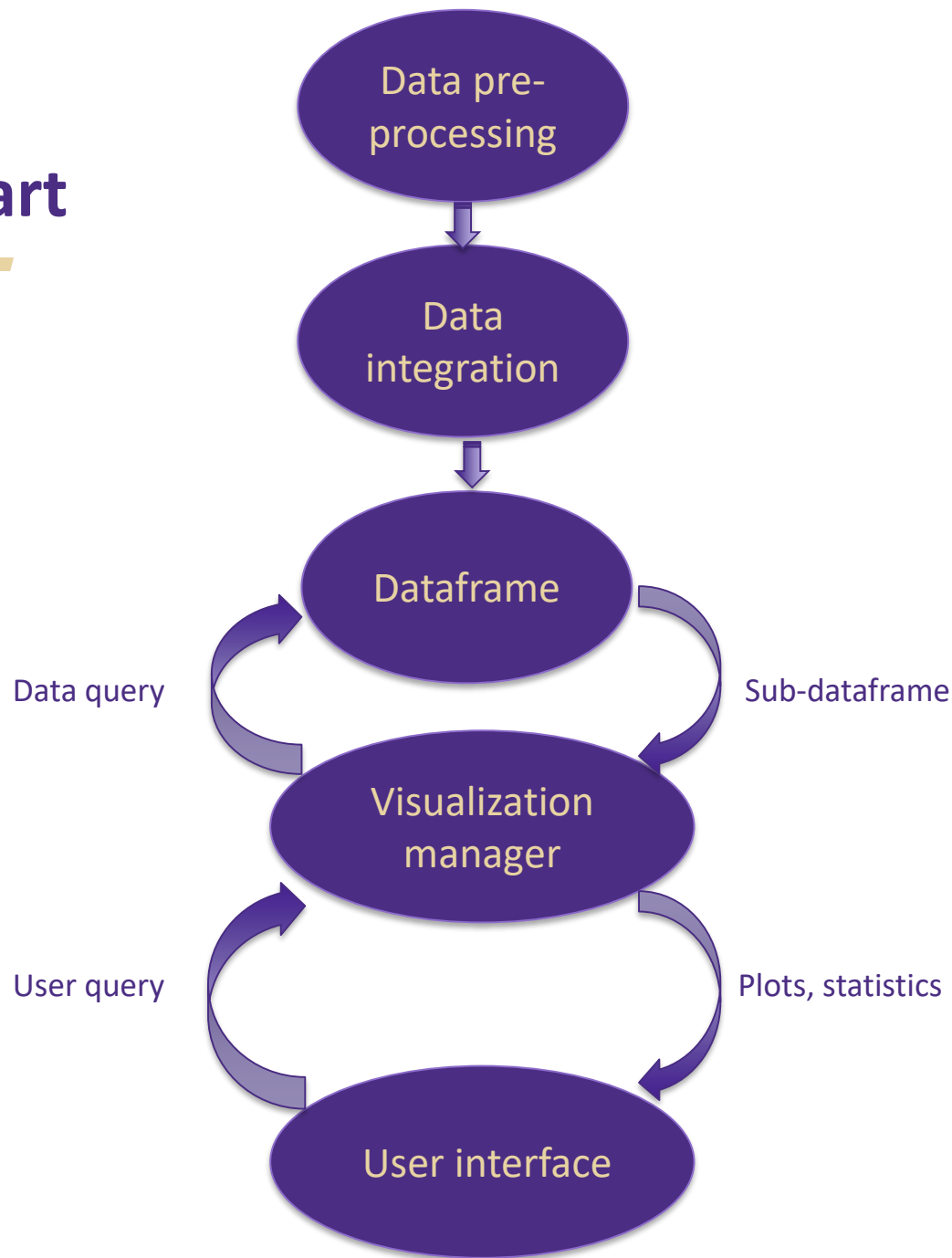
---

- > **Data cleaning and merging: contains functions to merge items.csv and reviews.csv based on product id.**
  - Inputs: Two dataframes of raw data
  - Outputs: A merged and clean dataframe containing both item-level and user-level information.
- > **Visualization manager: extract and gather sales amount, ratings, review, etc. from the merged dataframe.**
  - Inputs: A clean and merged dataframe, and a specific query.
  - Outputs: Visualization of the query (e.g. histogram, pie chart)
- > **Web manager: build an web interface which manipulates HTML DOM object, enable users to interact with our system.**
  - Inputs: Get the values from the drop down box
  - Outputs: Plots from visualization manager



# Flow chart

---



# Project Structure

---

```
|-- AMZ-Cellphone-Review
|   |   | interface.py
|   |   | visualization_functions.py
|   |   | data
|   |   |     | items.csv
|   |   |     | reviews.csv
|   |   | test
|   |   |     | test_visualization_functions.py
|-- Docs
|   | Component_specification.md
|   | Functional_specification.md
|   | Homework\ 5_\ Project\ Design.docx
|   | technologies_reviews.pptx
|-- images
|   | images1.png
|   | images2.png
|-- LICENSE
|-- README.md
|-- requirements.txt
|-- setup.py
|-- .travis.yml
|-- .gitignore
```



# Lessons learned and future work

---

- > **Components that a software should include**
  - Test module : edge case, unit test, expectation of input and output, continuous integration
  - Complete documentation
  - Structure of Python package
- > **Future work**
  - Sentiment analysis for reviews
  - Regularly update data

