

Cell Phone Reviews

PRI · Information Processing and Retrieval, 2021/22

M.EIC · Master in Informatics and Computing Engineering

T4G5 · Milestone #1

André Gomes - up201806224

André Nascimento - up201806461

Catarina Fernandes - up201806854

Downloaded 2019 dataset from Kaggle ([Amazon Cell Phones Reviews](#))

Scraped new 2021 information, adapting the scraper used for the kaggle dataset

Items

- csv files, 10 columns
- 720 entries in 2019 dataset
- 823 entries in 2021 dataset

Reviews

- csv files, 8 columns
- 67986 entries in 2019 dataset
- 48002 entries in 2021 dataset

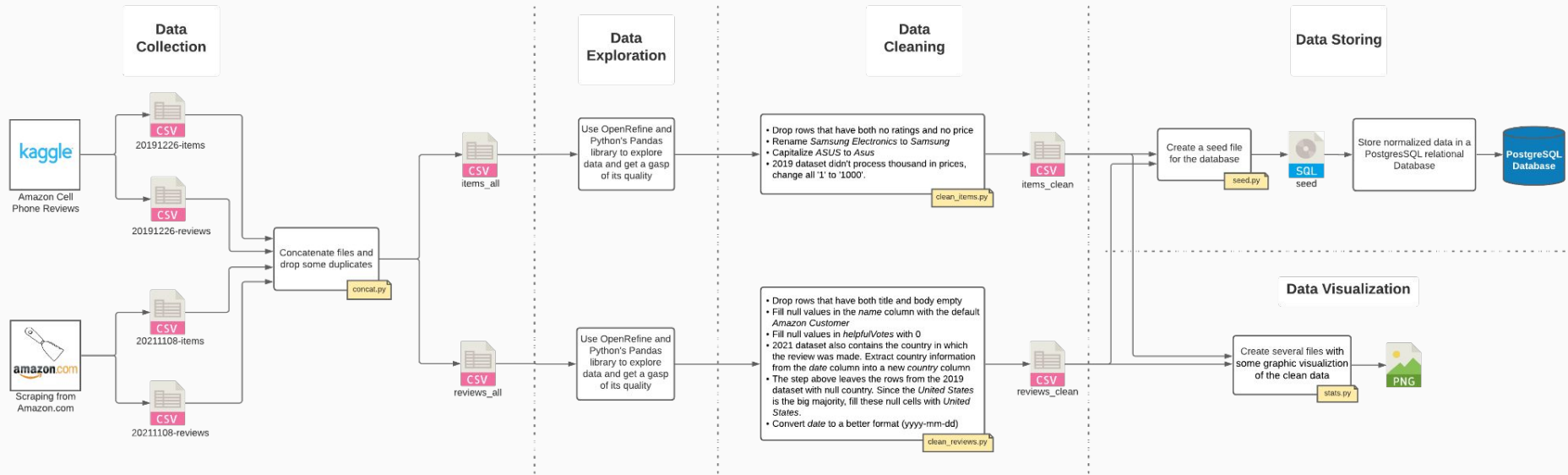
Items Dataset

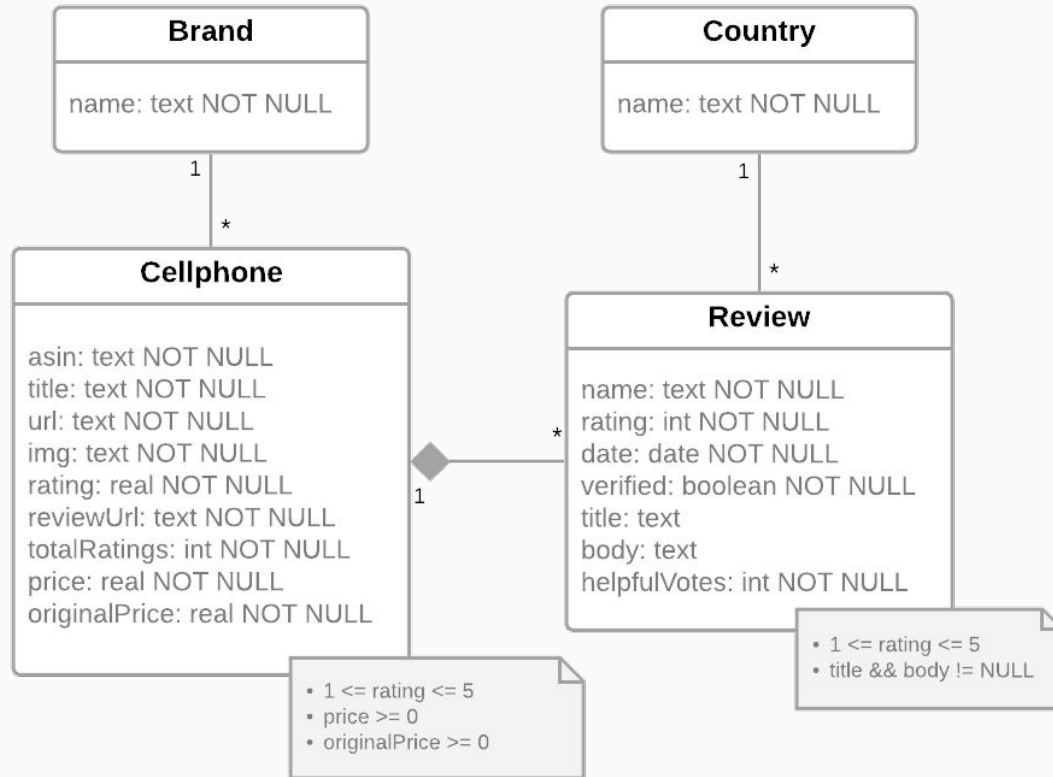
1	asin	brand	title	url	image	rating	reviewUrl	totalReviews	price	originalPrice
2	B0000SX2UC	Nokia	Dual-Band / T	https://wv	https://m.n	3	https://wv	14	0	0
3	B0009N5L7K	Motorola	Motorola I265	https://wv	https://m.n	3	https://wv	7	49.95	0
4	B000SKTZ0S	Motorola	MOTOROLA C	https://wv	https://m.n	2.7	https://wv	22	99.99	0
5	B001AO4OUC	Motorola	Motorola i335	https://wv	https://m.n	3.3	https://wv	21	0	0
6	B001DCJAJG	Motorola	Motorola V36	https://wv	https://m.n	3.1	https://wv	12	149.99	0
7	B001GQ3DJM	Nokia	Nokia 1680 Bl	https://wv	https://m.n	2.7	https://wv	3	0	0
8	B0029F2O3A	Samsung	Samsung Alias	https://wv	https://m.n	3	https://wv	29	0	0
9	B002AS9WEA	Samsung	Samsung a167	https://wv	https://m.n	2.9	https://wv	18	0	0
10	B002UHS0UI	Motorola	Verizon Wirele	https://wv	https://m.n	2.7	https://wv	208	0	0
11	B002WTC1NG	Motorola	Motorola Bar	https://wv	https://m.n	3.6	https://wv	352	139.99	0

Review Dataset

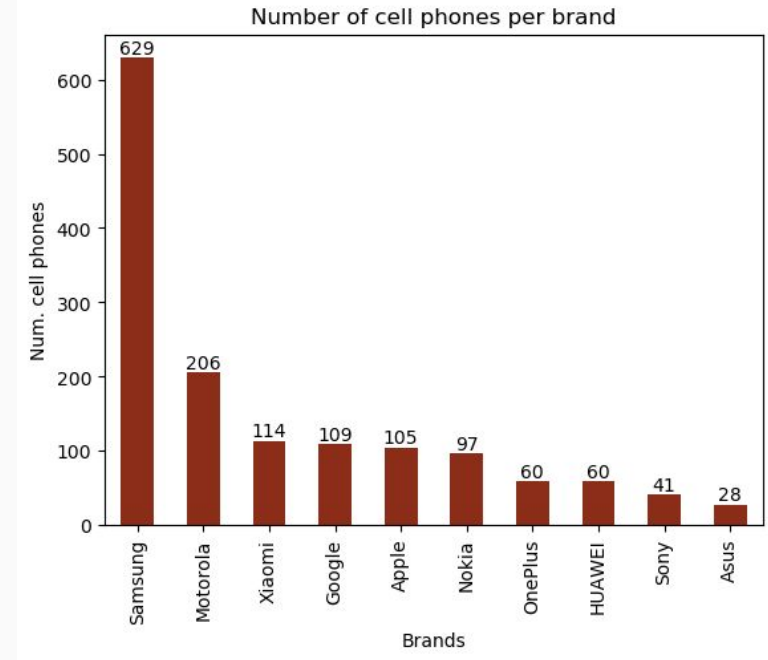
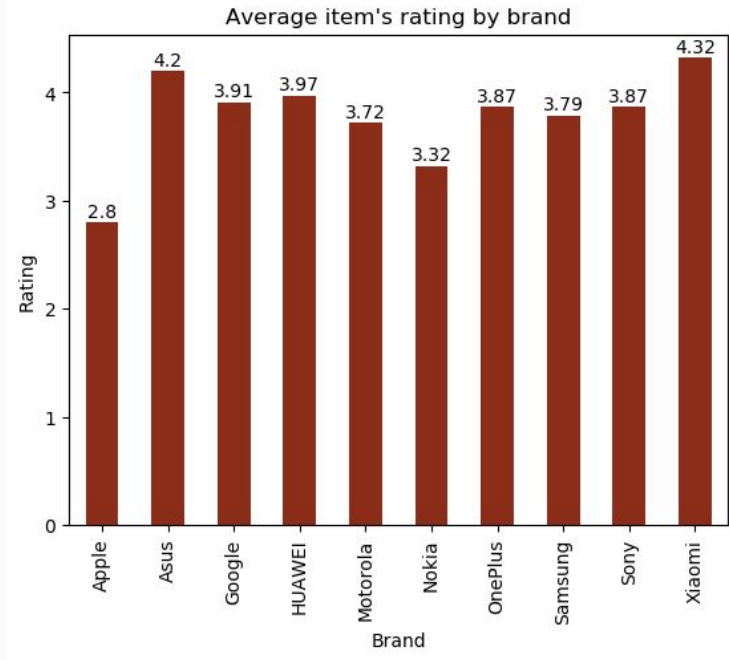
1	asin	name	rating	date	verified	title	body	helpfulVotes
2	B0000SX2UC	Janet	3	October 11, 2005	FALSE	Def not best, but	I had the Samsun	1
3	B0000SX2UC	Luke Wyatt	1	January 7, 2004	FALSE	Text Messaging D	Due to a software	17
4	B0000SX2UC	Brooke	5	December 30, 2003	FALSE	Love This Phone	This is a great, rel	5
5	B0000SX2UC	amy m. tea	3	March 18, 2004	FALSE	Love the Phone, E	I love the phone a	1
6	B0000SX2UC	tristazbimr	4	August 28, 2005	FALSE	Great phone servi	The phone has be	1
7	B0000SX2UC	J. White	4	September 25, 2005	FALSE	Worked great for	Hello, I have this	
8	B0000SX2UC	the cell ph	5	April 16, 2004	FALSE	Wanna cool Noki	Cool. Cheap. Col	2
9	B0000SX2UC	Matt	4	April 3, 2004	FALSE	Problem with 358	The 3599i is over	2
10	B0000SX2UC	Charles Co	5	November 24, 2003	FALSE	cool phone!!!!!!!	I've never owned	7
11	B0000SX2UC	Amazon C	3	February 2, 2004	FALSE	Pissed off-a little	ok well im in scho	3

Pipeline





Characterize the dataset - Items



U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO



Future work - Queries



- Search Cell Phones
 - By brand, title, rating, price
- Search Reviews
 - By reviewer's name, title, body, country
- Most/Least liked brands
- Most/Least expensive cell phones
- Most used words/expressions on reviews titles and bodies
- Most/Least Helpful Reviews