PROJECT 2 – ETL

ASX - 200

1st of August 2022

Group 9;

David Salim
Christopher Mai
Johan Ehrhardt
Krystal Enweya

CONTENTS;

1. Project aim;

The aim of our project was to collect the data regarding the ASX 200. From this collected data the aim was to create data bases and join tables to provide a relevant and effective summary of the ASX 200

2. Extraction;

Extraction of the data used within this project can be divided based on content of tables;

Table 1, 'ASX 200 Companies'

- Retrieved from the following website; https://www.marketindex.com.au/asx200
- data regarding the ASX 200
- Base data from web therefore using web scrapping
- Splinter and google chrome driver manager used to interact with the web
- Beautiful soup and for loop used to extract data from the table
- Pandas used to create data frames from the extracted data

Code	Company	Price	Chg	% Chg	High	Low	Volume	Mkt Cap	1 Year
A200	Betashares Australia 200 ETF	$116.21	+0.71	+0.62%	$116.30	$115.59	46,534	$2.2B	-6.97%
A2M	The a2 Milk Company Ltd	$4.54	0.00	0.00	$4.60	$4.49	2,382,120	$3.4B	-22.53%
AAA	Betashares Australian High Interest Cash ETF	$50.08	-0.05	-0.10%	$50.08	$50.07	43,341	$1.8B	-0.02%
ABP	Abacus Property Group	$2.92	-0.01	-0.34%	$2.96	$2.89	1,158,318	$2.6B	-7.89%
AFI	Australian Foundation Investment Company Ltd	$8.02	+0.02	+0.25%	$8.14	$7.99	434,943	$9.9B	-4.75%
AGL	AGL Energy Ltd	$8.45	+0.07	+0.84%	$8.45	$8.325	1,110,766	$5.7B	+16.87%
AIA	Auckland International Airport Ltd	$6.71	+0.04	+0.60%	$6.74	$6.65	471,451	$9.9B	-1.32%
AIZ	Air New Zealand Ltd	$0.565	+0.02	+3.67%	$0.565	$0.545	719,103	$1.9B	-60.21%
AKE	Allkem Ltd	$11.79	+0.51	+4.52%	$11.92	$11.30	4,585,892	$7.5B	+42.91%
ALD	Ampol Ltd	$34.01	+0.51	+1.52%	$34.14	$33.50	680,050	$8.1B	+20.39%
ALL	Aristocrat Leisure Ltd	$35.21	-0.09	-0.26%	$35.55	$35.10	1,910,979	$23B	-15.10%
ALQ	Als Ltd	$11.66	+0.02	+0.17%	$11.75	$11.53	910,430	$5.6B	-8.48%
ALU	Altium Ltd	$30.64	-0.53	-1.70%	$31.63	$30.56	284,540	$4B	-9.62%
ALX	Atlas Arteria	$7.76	+0.01	+0.13%	$7.96	$7.71	5,021,120	$7.4B	+23.76%
AMC	Amcor Plc	$18.58	+0.09	+0.49%	$18.58	$18.43	1,896,189	$13B	+17.97%
AMP	AMP Ltd	$1.105	+0.02	+1.84%	$1.11	$1.085	9,524,894	$3.6B	+6.25%

Figure 1. Raw data, from web

Table 2, ‘ASX Companies’

- Retrieved from the following website; https://www.listcorp.com/asx/
- Data regarding ASX Companies
- CSV file download
- Pandas to extract data
- Pandas used to create data frames from the extracted data

	A	B	C	D	E	F	G	H	I	J
1	Code	Company	Link	Market Ca	Last trade	Change	%Change	Sector		
2	ASX:BHP	BHP Grouf	https://wv	193988000	38.68	0	0	Materials		
3	ASX:CBA	Commonw	https://wv	169967000	100.77	0	0	Financials		
4	ASX:CSL	CSL Limite	https://wv	140369000	289.84	0	0	Health Care		
5	ASX:NAB	National A	https://wv	96384200(	30.6	0	0	Financials		
6	ASX:WBC	Westpac B	https://wv	75029200(	21.51	0	0	Financials		
7	ASX:MQG	Macquarie	https://wv	69136800(	181.13	0	0	Financials		
8	ASX:ANZ	ANZ Banki	https://wv	63624000(	22.9	0	0	Financials		
9	ASX:WDS	Woodside	https://wv	60703000(	31.98	0	0	Energy		
10	ASX:FMG	Fortescue	https://wv	57576600(	18.34	0	0	Materials		
11	ASX:WES	Wesfarme	https://wv	52496800(	46.63	0	0	Consumer Discretionary		
12	ASX:WOW	Woolwort	https://wv	45448500(	37.52	0	0	Consumer Staples		
13	ASX:TLS	Telstra Lim	https://wv	45062300(	3.89	0	0	Communication Services		
14	ASX:TCL	Transurbar	https://wv	44095500(	14.51	0	0	Industrials		
15	ASX:GMG	Goodman	https://wv	36972100(	20.7	0	0	Real Estate		
16	ASX:RIO	Rio Tinto (	https://wv	36267800(	97.83	0	0	Materials		
17	ASX:COL	Coles Grou	https://wv	24908400(	18.75	0	0	Consumer Staples		
18	ASX:STO	Santos Lim	https://wv	24293500(	7.3	0	0	Energy		
19	ASX:ALL	Aristocrat	https://wv	24009400(	35.3	0	0	Consumer Discretionary		
20	ASX:QBE	QBE Insura	https://wv	17606300(	11.53	0	0	Financials		
21	ASX:S32	South32 Li	https://wv	17495500(	3.81	0	0	Materials		
22	ASX:ASX	ASX Limite	https://wv	16844700(	88.26	0	0	Financials		
23	ASX:NCM	Newcrest	https://wv	16621000(	19.3	0	0	Materials		
24	ASX:REA	REA Grouf	https://wv	16320400(	125.06	0	0	Communication Services		
25	ASX:SHL	Sonic Heal	https://wv	16087900(	34.27	0	0	Health Care		
26	ASX:WTC	WiseTech	https://wv	15986500(	50.1	0	0	Information Technology		
27	ASX:RHC	Ramsay He	https://wv	15985100(	70.2	0	0	Health Care		
28	ASX:BXB	Brambles l	https://wv	15795500(	11.45	0	0	Industrials		
29	ASX:JHX	James Har	https://wv	15411900(	35.2	0	0	Materials		

companies-list

Figure 2. Raw data, as CSV file

3. Transformation;

The following data cleaning was performed in order to transform the data into concise and relevant pieces information;

3.1 Removing columns not relevant to our aim including;

From Table 1;
- High price
- Low price
- Volume
- Market cap

From Table 2;
- Link
- Last trade
- change
- % change

3.2 Dropping duplicates to ensure high data validity

3.3 Using functions such as;
- 'Fill na' for table 1 data to ensure our sample size remains (200)
- 'Drop na' for table 2 data

3.4 Formatting data to present aesthetically;
- Removing '$' from each row
- Removing 'ASX:' prefix from row company titles by renaming content of columns
- Sort data to allow for more efficient tracking

3.5 Ensure datatypes are appropriate for each variable in each column
- E.g. transforming 'Object' -> 'Float'

```
# show data
print(len(transformed_asx200_df))
transformed_asx200_df.head()
```

✓ 0.4s Python

```
200
```

	company_code	company_name	price_29jul22	change	percent_change	one_year_percent_change
0	A200	Betashares Australia 200 ETF	116.21	0.71	0.62	-6.97
1	A2M	The a2 Milk Company Ltd	4.54	0.00	0.00	-22.53
2	AAA	Betashares Australian High Interest Cash ETF	50.08	-0.05	-0.10	-0.02
3	ABP	Abacus Property Group	2.92	-0.01	-0.34	-7.89
4	AFI	Australian Foundation Investment Company Ltd	8.02	0.02	0.25	-4.75

Figure 3. Transformed 'ASX 200' data from Table 1

```
# show data
print(len(transformed_asx_company_df))
transformed_asx_company_df.head()
```

✓ 0.2s Python

2217

	company_code	company_name	market_cap	sector
0	14D	1414 Degrees Limited (ASX:14D)	21208500	Industrials
1	1AD	AdAlta Limited (ASX:1AD)	15709200	Health Care
2	1AE	Aurora Energy Metals Limited (ASX:1AE)	26270400	Materials
3	1AG	Alterra Limited (ASX:1AG)	11128800	Consumer Staples
4	1MC	Morella Corporation Limited (ASX:1MC)	93171900	Materials

Figure 4. Transformed 'ASX Company' data from Table 2

4. Loading;

4.1 Create SQL database 'asx_db' in pg postgress

- Create two tables to insert values into
  - Asx 200 table
  - Asx_companies table

4.2 Create connection in jupyter notebook to postgress to allow loading of data

4.3 Load data frames created into the sql database

```
1 select * from asx_companies
```

Data output | Messages | Notifications

	company_code [PK] character varying (10)	company_name character varying (100)	market_cap numeric	sector character varying (100)
1	14D	1414 Degrees Limited (ASX:14D)	21208500	Industrials
2	1AD	AdAlta Limited (ASX:1AD)	15709200	Health Care
3	1AE	Aurora Energy Metals Limited (ASX:1AE)	26270400	Materials
4	1AG	Alterra Limited (ASX:1AG)	11128800	Consumer Staples
5	1MC	Morella Corporation Limited (ASX:1MC)	93171900	Materials
6	1ST	1st Group Limited (ASX:1ST)	9017220	Health Care
7	1VG	Victory Goldfields Limited (ASX:1VG)	7114760	Materials
8	29M	29Metals Limited (ASX:29M)	714759000	Materials

Total rows: 1000 of 2217 | Query complete 00:00:00.205 | Ln 1, Col 28

Figure 5. Asx_companies table

```
1 select * from asx_200;
```

Data output | Messages | Notifications

	company_code [PK] character varying (10)	company_name character varying (100)	price_29jul22 numeric	change numeric	percent_change numeric	one_year_percent_change numeric
1	A200	Betashares Australia 200 ETF	116.21	0.71	0.62	-6.97
2	A2M	The a2 Milk Company Ltd	4.54	0.0	0.0	-22.53
3	AAA	Betashares Australian High Interest Cash ETF	50.08	-0.05	-0.1	-0.02
4	ABP	Abacus Property Group	2.92	-0.01	-0.34	-7.89
5	AFI	Australian Foundation Investment Company Ltd	8.02	0.02	0.25	-4.75
6	AGL	AGL Energy Ltd	8.45	0.07	0.84	16.87
7	AIA	Auckland International Airport Ltd	6.71	0.04	0.6	-1.32
8	AIZ	Air New Zealand Ltd	0.565	0.02	3.67	-60.21

Total rows: 200 of 200 | Query complete 00:00:00.124 | Ln 1, Col 23

Figure 6. Asx_200 table

4.4 Joining tables on 'company_code'

4.5 Create a ‘view’ for the joined table ‘asx_200_data’ to prevent manual joining at each moment that joint data is to be analysed

```
1  select * from asx_200_data
```

Data output | Messages | Notifications

	company_code character varying (10)	company_name character varying (100)	price_29jul22 numeric	percent_change numeric	one_year_percent_change numeric	sector character varying (100)	market_cap numeric
1	A200	Betashares Australia 200 ETF	116.21	0.62	-6.97	Financials	2170880000
2	A2M	The a2 Milk Company Ltd	4.54	0.0	-22.53	Consumer Staples	3272090000
3	ABP	Abacus Property Group	2.92	-0.34	-7.89	Real Estate	2570200000
4	AFI	Australian Foundation Investment Company Ltd	8.02	0.25	-4.75	Financials	9839250000
5	AGL	AGL Energy Ltd	8.45	0.84	16.87	Utilities	5570350000
6	AIA	Auckland International Airport Ltd	6.71	0.6	-1.32	Industrials	9674660000
7	AIZ	Air New Zealand Ltd	0.565	3.67	-60.21	Industrials	1835810000
8	AKE	Allkem Ltd	11.79	4.52	42.91	Materials	7090760000

Total rows: 200 of 200 | Query complete 00:00:00.135 | Ln 1, Col 27

Figure 7. Asx_200_data , joint table

5. Schema;

project_no_2.sql

```sql
-- create asx_200 table
CREATE TABLE asx_200 (
company_code VARCHAR(10) NOT NULL,
company_name VARCHAR(100) NOT NULL,
price_29jul22 DEC,
change DEC,
percent_change DEC,
one_year_percent_change DEC,
PRIMARY KEY (company_Code)
);

-- create asx_companies table
CREATE TABLE asx_companies (
company_code VARCHAR(10) NOT NULL,
company_name VARCHAR(100) NOT NULL,
market_cap DEC,
sector VARCHAR(100) NOT NULL,
PRIMARY KEY (company_code)
);

-- Join tables on company_code
SELECT a200.company_code, a200.company_name, a200.price_29jul22, a200.percent_change, a200.one_year_percent_change, a.sector , a.market_cap
FROM asx_200 a200
LEFT JOIN asx_companies a
ON a200.company_code = a.company_code;

-- create view
CREATE VIEW asx_200_data AS
SELECT a200.company_code, a200.company_name, a200.price_29jul22, a200.percent_change, a200.one_year_percent_change, a.sector , a.market_cap
FROM asx_200 a200
LEFT JOIN asx_companies a
ON a200.company_code = a.company_code;
```

Figure 8. Schema

6. ERD Diagram;