

Moscholios Filippou-Michal, F3352116

Queries on Monet DB using Python

This deliverable contains explanation for each query implementing in Monet DB and the corresponding outputs for each query. For each query the rationale was to pass the corresponding content from the Pyspark code into the Monet DB context. In more details, the preprocessing for the udfs was the same, but this time for each listing after the preprocessing, the results stored into a list. This procedure taking place until all listings filtered and stored into a list. Therefore, the outcome of the query has the desired filtering format. **Also, to make the best use of the reserved space, for the following udf queries, the explanation is the same as presented in the attached deliverable of F3352116_Pyspark-Report-and-Code, regarding the usage of split function.** Moreover, including some statistics for some queries such as the number of listings after filtering.

First, in order to load the data properly, we construct a loader function named 'loader_dataset()' written in language Python. Inside of the function we start by importing the Pandas library for better manipulation. We read the csv file from the corresponding path and starting by parsing the file. In the sequel, using loc function from Pandas, we emit each column from the data frame. Hence, with the feature[0] we store all the listings from the first column which is the 'title' and named the corresponding column with the same name. This procedure continues until every column of the data frame emitted. The last part of this processing is to create a table called 'dataset' which includes the content of the loader_dataset(). At the next depiction **Table 1**, we project the output of our table called dataset.

OUTPUT

```
sql>SELECT * FROM dataset
more>LIMIT 10
more>
```

title	address	city	state	postal_code	price	facts and features	real estate provider	url
Condo for sale	rville-MA-02145/2077576269_zpid/	null	Somerville	MA	null	\$342,000 2 bds, 1.0 ba ,705 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-204-Some
Condo for sale	Boston-MA-02116/2124477129_zpid/	null	Boston	MA	null	\$1,700,000 2 bds, 2.0 ba ,1228 sqft	Century 21 North East	https://www.zillow.com/homedetails/6-Edgerly-Pl-APT-401
Condo for sale	St-209-Boston-MA-02118/59211169_zpid/	null	Boston	MA	null	\$336,500 1 bds, 1.0 ba ,1000 sqft	Maloney Properties, Inc.	https://www.zillow.com/homedetails/1180-1200-Washington
House for sale	MA-02118/103861295_zpid/	null	Boston	MA	null	\$9,950,000 4 bds, 7.0 ba ,6836 sqft	Campan & Company Fine Homes Real Estate	https://www.zillow.com/homedetails/5-Union-Park-Boston-
Condo for sale	ton-MA-02128/2069976926_zpid/	null	Boston	MA	null	\$479,000 2 bds, 3.0 ba ,1000 sqft	Berkshire Hathaway HomeServices Commonwealth Real Estate	https://www.zillow.com/homedetails/213-Trenton-St-A-Bos
House for sale	oston-MA-02128/59123230_zpid/	null	East Boston	MA	null	\$899,000 3 bds, 3.0 ba ,2313 sqft	Berkshire Hathaway HomeServices Warren Residential	https://www.zillow.com/homedetails/194-London-St-East-B
Condo for sale	rville-MA-02145/2077576284_zpid/	null	Somerville	MA	null	\$397,300 2 bds, 1.0 ba ,780 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-102-Some
Condo for sale	ton-MA-02127/59148399_zpid/	null	South Boston	MA	null	\$619,900 2 bds, 1.0 ba ,856 sqft	Bento Real Estate Group, Inc.	https://www.zillow.com/homedetails/144-O-St-2-South-Bos
Condo for sale	APT-106-Boston-MA-02116/59177162_zpid/	null	Boston	MA	null	\$850,000 1 bds, 1.0 ba ,675 sqft	Engel & Volkers Boston	https://www.zillow.com/homedetails/280-Commonwealth-Ave
Condo for sale	n-MA-02114/59174070_zpid/	null	Boston	MA	null	\$649,900 2 bds, 1.0 ba ,511 sqft	Leading Edge Real Estate	https://www.zillow.com/homedetails/55-Garden-St-9-Bosto

```
10 tuples
```

Table 1

Query 1

Mentioning here that before the splitting phase taking place, we construct a list called *bedrooms* to append each listing after filtering, having at last the desire filtered format. The functions name is *retrieve_bedrooms*. The output below **Figure 1** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_bedrooms(dataset."facts and features") AS bedrooms from dataset limit 10;
+-----+
| bedrooms |
+-----+
| 2 |
| 2 |
| 1 |
| 4 |
| 2 |
| 3 |
| 2 |
| 2 |
| 1 |
| 2 |
+-----+
10 tuples
```

Figure 1

Query 2

Mentioning here that for each listing after filtering the outcome stored into a list called *bathrooms*. The functions name is *retrieve_bathrooms*. The output below **Figure 2** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_bathrooms(dataset."facts and features") AS bathrooms from dataset limit 10;
+-----+
| bathrooms |
+-----+
| 1 |
| 2 |
| 1 |
| 7 |
| 3 |
| 3 |
| 1 |
| 1 |
| 1 |
| 1 |
+-----+
10 tuples
```

Figure 2

Query 3

Mentioning here that for each listing after filtering the outcome stored into a list called *sqft*. The functions name is *retrieve_sq_ft*. The output below **Figure 3** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_sq_ft(dataset."facts and features") AS SQFT from dataset limit 10;
+-----+
| sqft |
+-----+
| 705   |
| 1228  |
| 1000  |
| 6836  |
| 1000  |
| 2313  |
| 780   |
| 856   |
| 675   |
| 511   |
+-----+
10 tuples
```

Figure 3

Query 4

Mentioning here that for each listing after filtering the outcome stored into a list called *type*. The functions name is *retrieve_type*. The output below **Figure 4** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_type(dataset."title") AS TYPE from dataset limit 10;
+-----+
| type |
+-----+
| Condo |
| Condo |
| Condo |
| House |
| Condo |
| House |
| Condo |
| Condo |
| Condo |
| Condo |
+-----+
10 tuples
```

Figure 4

Query 5

Mentioning here that for each listing after filtering the outcome stored into a list called *offer*. The functions name is *retrieve_offer*. The output below **Figure 5** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_offer(dataset."title") as OFFER from dataset limit 10;
+-----+
| offer |
+-----+
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
| sale  |
+-----+
10 tuples
```

Figure 5

Query 6

For this query we filter out the listings which are not for sale. First, we use the udf *retrieve_offer* constructed at the **Query 5**, and we filter the resulting column, using the Like equal “%sale%” to get only the listings that contain sale, therefore filtering out the listings that are not for sale. The results presented on **Figure 6**. Afterwards we provide some statistics presenting the number of listings after the filtering, in **Table 2**.

OUTPUT

```
sql>SELECT *
FROM dataset
WHERE retrieve_offer(dataset."title") LIKE '%sale%' limit 10;
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| title | address | city | state | postal_code | price | facts and features | real estate provider | url |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Condo for sale | | Somerville | MA | null | $342,000 | 2 bds, 1.0 ba ,705 sqft | William Raveis R.E. & Home Services | https://www.zillow.com/homedetails/163-Glen-St-204-Some- |
| ville-MA-02145/2077576269_zpid/ | | Boston | MA | null | $1,700,000 | 2 bds, 2.0 ba ,1228 sqft | Century 21 North East | https://www.zillow.com/homedetails/6-Edgerly-Pk-APT-401- |
| Boston-MA-02116/214477129_zpid/ | | Boston | MA | null | $336,500 | 1 bds, 1.0 ba ,1000 sqft | Maloney Properties, Inc. | https://www.zillow.com/homedetails/1100-1200-Washington- |
| ST-209-Boston-MA-02118/59211169_zpid/ | | Boston | MA | null | $9,950,000 | 4 bds, 7.0 ba ,6836 sqft | Campton & Company Fine Homes Real Estate | https://www.zillow.com/homedetails/5-Union-Park-Boston- |
| House for sale | | Boston | MA | null | $479,000 | 2 bds, 3.0 ba ,1000 sqft | Berkshire Hathaway HomeServices Commonwealth Real Estate | https://www.zillow.com/homedetails/213-Trenton-St-A-Bos- |
| MA-02118/103061295_zpid/ | | East Boston | MA | null | $899,000 | 3 bds, 3.0 ba ,2313 sqft | Berkshire Hathaway HomeServices Warren Residential | https://www.zillow.com/homedetails/194-London-St-East-B- |
| Condo for sale | | Somerville | MA | null | $397,300 | 2 bds, 1.0 ba ,780 sqft | William Raveis R.E. & Home Services | https://www.zillow.com/homedetails/163-Glen-St-102-Some- |
| ville-MA-02145/2077576284_zpid/ | | South Boston | MA | null | $619,900 | 2 bds, 1.0 ba ,856 sqft | Bento Real Estate Group, Inc. | https://www.zillow.com/homedetails/144-0-St-2-South-Bos- |
| Condo for sale | | Boston | MA | null | $850,000 | 1 bds, 1.0 ba ,675 sqft | Engel & Volkers Boston | https://www.zillow.com/homedetails/280-Commonwealth-Ave- |
| ton-MA-02117/59148399_zpid/ | | Boston | MA | null | $649,900 | 2 bds, 1.0 ba ,511 sqft | Leading Edge Real Estate | https://www.zillow.com/homedetails/55-Garden-St-9-Bosto- |
| APT-106-Boston-MA-02116/59177162_zpid/ | | | | | | | | |
| MA-02114/59174070_zpid/ | | | | | | | | |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+
10 tuples
```

Figure 6

STATISTICS OUTPUT

```
sql>SELECT COUNT(*) AS number_of_listins_which_are_for_sale
more>FROM dataset
more>WHERE retrieve_offer(dataset."title") LIKE '%sale%' limit 10;
+-----+
| number_of_listins_which_are_for_sale |
+-----+
| 100017 |
+-----+
1 tuple
```

Table 2

Query 7

Mentioning here that for each listing after filtering the outcome stored into a list called *price_filtered*. The functions name is *retrieve_filter_price*. The output below **Figure 7** showcasing the first 10 results after the processing.

OUTPUT

```
sql>select retrieve_filter_price(dataset."price") as price from dataset limit 10;
+-----+
| price |
+-----+
| 342000 |
| 1700000 |
| 336500 |
| 9950000 |
| 479000 |
| 899000 |
| 397300 |
| 619900 |
| 850000 |
| 649900 |
+-----+
10 tuples
```

Figure 7

Query 8

For this query we filter out the listings which are not for sale. First, we use the udf *retrieve_bedrooms* constructed at the **Query 1**, and we filter the resulting column, using less than “11” bedrooms to get only the listings that contain less than 11 bedrooms, therefore filtering out the listings that have more than 11 bedrooms. The results presented on **Figure 8**. Afterwards we provide some statistics presenting the number of listings after the filtering, in **Table 3**.

OUTPUT

```
sql>SELECT *
more>FROM dataset
more>WHERE retrieve_bedrooms(dataset."facts and features") < 11 limit 10;
```

title	address	city	state	postal_code	price	facts and features	real estate provider	url
Condo for sale		null	Somerville	MA	null	\$342,000 2 bds, 1.0 ba ,705 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-204-Some
Condo for sale		null	Boston	MA	null	\$1,700,000 2 bds, 2.0 ba ,1228 sqft	Century 21 North East	https://www.zillow.com/homedetails/6-Edgerly-Pl-APT-401
Condo for sale		null	Boston	MA	null	\$136,500 1 bds, 1.0 ba ,1000 sqft	Maloney Properties, Inc.	https://www.zillow.com/homedetails/1180-1200-Washington
House for sale		null	Boston	MA	null	\$9,950,000 4 bds, 7.0 ba ,6836 sqft	Campan & Company Fine Homes Real Estate	https://www.zillow.com/homedetails/5-Union-Park-Boston-
House for sale		null	Boston	MA	null	\$479,000 2 bds, 3.0 ba ,1000 sqft	Berkshire Hathaway HomeServices Commonwealth Real Estate	https://www.zillow.com/homedetails/213-Trenton-St-A-Bos
House for sale		null	East Boston	MA	null	\$899,000 3 bds, 3.0 ba ,2313 sqft	Berkshire Hathaway HomeServices Warren Residential	https://www.zillow.com/homedetails/194-London-St-East-B
Condo for sale		null	Somerville	MA	null	\$397,300 2 bds, 1.0 ba ,780 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-102-Some
Condo for sale		null	South Boston	MA	null	\$619,900 2 bds, 1.0 ba ,856 sqft	Bento Real Estate Group, Inc.	https://www.zillow.com/homedetails/144-O-St-2-South-Bos
Condo for sale		null	Boston	MA	null	\$850,000 1 bds, 1.0 ba ,675 sqft	Engel & Volkers Boston	https://www.zillow.com/homedetails/280-Commonwealth-Ave
Condo for sale		null	Boston	MA	null	\$649,900 2 bds, 1.0 ba ,511 sqft	Leading Edge Real Estate	https://www.zillow.com/homedetails/55-Garden-St-9-Bosto

10 tuples

Figure 8

STATISTICS OUTPUT

```
sql>SELECT COUNT(*) AS number_of_listings_less_than_11_bedrooms
more>FROM dataset
more>WHERE retrieve_bedrooms(dataset."facts and features") < 11 limit 10;
```

number_of_listings_less_than_11_bedrooms
99663

1 tuple

Table 3

Query 9

For this query we filter out the listings which are not for sale. First, we use the udf *retrieve_filter_price* constructed at the **Query 7**, and we filter the resulting column, using less than “<” 2m and more than “>” 100k in order to get only the listings that belong to this range, therefore filtering out the listings that are extremely expensive and cheap meaning more than 2m and less than 100k. The results presented on **Figure 9**. Afterwards we provide some statistics presenting the number of listings after the filtering, in **Table 4**.

OUTPUT

```
sql>SELECT *
more>FROM dataset
more>WHERE retrieve_filter_price(dataset."price") < 2000000 AND retrieve_filter_price(dataset."price") > 100000 limit 10;
```

title	address	city	state	postal_code	price	facts and features	real estate provider	url
Condo for sale		null	Somerville	MA	null	\$342,000 2 bds, 1.0 ba ,705 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-204-Some
Condo for sale		null	Boston	MA	null	\$1,700,000 2 bds, 2.0 ba ,1228 sqft	Century 21 North East	https://www.zillow.com/homedetails/6-Edgerly-Pl-APT-401
Condo for sale		null	Boston	MA	null	\$136,500 1 bds, 1.0 ba ,1000 sqft	Maloney Properties, Inc.	https://www.zillow.com/homedetails/1180-1200-Washington
House for sale		null	Boston	MA	null	\$9,950,000 4 bds, 7.0 ba ,6836 sqft	Campan & Company Fine Homes Real Estate	https://www.zillow.com/homedetails/5-Union-Park-Boston-
House for sale		null	Boston	MA	null	\$479,000 2 bds, 3.0 ba ,1000 sqft	Berkshire Hathaway HomeServices Commonwealth Real Estate	https://www.zillow.com/homedetails/213-Trenton-St-A-Bos
House for sale		null	East Boston	MA	null	\$899,000 3 bds, 3.0 ba ,2313 sqft	Berkshire Hathaway HomeServices Warren Residential	https://www.zillow.com/homedetails/194-London-St-East-B
Condo for sale		null	Somerville	MA	null	\$397,300 2 bds, 1.0 ba ,780 sqft	William Raveis R.E. & Home Services	https://www.zillow.com/homedetails/163-Glen-St-102-Some
Condo for sale		null	South Boston	MA	null	\$619,900 2 bds, 1.0 ba ,856 sqft	Bento Real Estate Group, Inc.	https://www.zillow.com/homedetails/144-O-St-2-South-Bos
Condo for sale		null	Boston	MA	null	\$850,000 1 bds, 1.0 ba ,675 sqft	Engel & Volkers Boston	https://www.zillow.com/homedetails/280-Commonwealth-Ave
Condo for sale		null	Boston	MA	null	\$649,900 2 bds, 1.0 ba ,511 sqft	Leading Edge Real Estate	https://www.zillow.com/homedetails/55-Garden-St-9-Bosto

10 tuples

Figure 9

STATISTICS OUTPUT

```
sql>SELECT COUNT(*) AS listings_with_price_from_2m_to_100k
more>FROM dataset
more>WHERE retrieve_filter_price(dataset."price") < 20000000 AND retrieve_filter_price(dataset."price") > 100000 limit 10;
+-----+
| listings_with_price_from_2m_to_100k |
+-----+
| 98838 |
+-----+
1 tuple
```

Table 4

Query 10

For this query we filter out the listings which are not for sale. First, we use the udf *retrieve_type* constructed at the **Query 4**, and we filter the resulting column, using the Like equal “%House%” to get only the listings that contain House, therefore filtering out the listings that are not Houses. The results presented on **Figure 10**. Afterwards we provide some statistics presenting the number of listings after the filtering, in **Table 5**.

OUTPUT

```
sql>SELECT *
more>FROM dataset
more>WHERE retrieve_type(dataset."title") LIKE '%House%' limit 10;
+-----+-----+-----+-----+-----+-----+-----+-----+
| title | address | city | state | postal_code | price | facts and features | real estate provider | url |
+-----+-----+-----+-----+-----+-----+-----+-----+
| House for sale | null | Boston | MA | null | $9,950,000 | 4 bds, 7.0 ba ,6836 sqft | Campion & Company Fine Homes Real Estate | https://www.zillow.com/homedetails/5-Union-Park-Boston-MA-02118/103861295_zpid/ |
| House for sale | null | East Boston | MA | null | $899,000 | 3 bds, 3.0 ba ,2313 sqft | Berkshire Hathaway HomeServices Warren Residential | https://www.zillow.com/homedetails/194-London-St-East-Boston-MA-02128/59123230_zpid/ |
| House for sale | null | Boston | MA | null | $1,200,000 | 2 bds, 1.0 ba ,1165 sqft | CL Properties | https://www.zillow.com/homedetails/44-Hull-St-Boston-MA-02113/59181015_zpid/ |
| House for sale | null | Boston | MA | null | $1,119,000 | 3 bds, 4.0 ba ,1680 sqft | All Star Realty, Inc. | https://www.zillow.com/homedetails/48-Allston-St-Boston-MA-02125/59099773_zpid/ |
| House for sale | null | South Boston | MA | null | $1,699,000 | 4 bds, 3.0 ba ,2043 sqft | Compass | https://www.zillow.com/homedetails/4-Linden-St-South-Boston-MA-02127/59150029_zpid/ |
| House for sale | null | Boston | MA | null | $589,000 | 2 bds, 2.0 ba ,1200 sqft | Gibson Sotheby's International Realty | https://www.zillow.com/homedetails/602-Saratoga-St-Boston-MA-02128/2060067352_zpid/ |
| House for sale | null | Boston | MA | null | $9,750,000 | 4 bds, 7.0 ba ,4754 sqft | MGS Group Real Estate LTD | https://www.zillow.com/homedetails/352-Marlborough-St-Boston-MA-02115/59169848_zpid/ |
| House for sale | null | Somerville | MA | null | $2,075,000 | 4 bds, 4.0 ba ,3121 sqft | Gibson Sotheby's International Realty | https://www.zillow.com/homedetails/16-Webster-St-Somerville-MA-02145/56334200_zpid/ |
| House for sale | null | Boston | MA | null | $3,200,000 | 3 bds, 3.0 ba ,2048 sqft | William Raveis R.E. & Home Services | https://www.zillow.com/homedetails/6-Strong-Pl-Boston-MA-02114/59173839_zpid/ |
| House for sale | null | South Boston | MA | null | $1,175,000 | 3 bds, 3.0 ba ,1867 sqft | Coldwell Banker Realty - South End | https://www.zillow.com/homedetails/23-Linden-St-South-Boston-MA-02127/59150016_zpid/ |
+-----+-----+-----+-----+-----+-----+-----+-----+
10 tuples
```

Figure 10

STATISTICS OUTPUT

```
sql>SELECT COUNT(*) AS number_of_listins_which_are_houses
more>FROM dataset
more>WHERE retrieve_type(dataset."title") LIKE '%House%' limit 10;
+-----+
| number_of_listins_which_are_houses |
+-----+
| 22554 |
+-----+
1 tuple
```

Table 5

Query 11

At the final query we must calculate the average price per sqft for house for sale and group them by the number of bedrooms. To achieve that we start by using the resulting filtering column from the **Query 1** which contain the number of bedrooms. We pass this attribute on the select statement to project it and on the group by. Moreover, we use the resulting filtering column from the **Query 7** which contain the filtered prices and divide them with the resulting filtering column from the **Query 3** which contain the filtered sqft. To conclude, at the select statement we took the AVG() of the division above in order to receive the average price per sqft. At the where statement we reference from the title column, for listings that are 'House for Sale' and constraint that we don't accept the sqft with 0 as value to avoid problems regarding the division. The above results depicted on **Figure 11**.

OUTPUT

```
sql>SELECT retrieve_bedrooms(dataset."facts and features") AS bedrooms, AVG(retrieve_filter_price(dataset."price")/retrieve_sq_ft(dataset."facts and features")) AS average_price_per_sqft from dataset
more>WHERE (title = 'House for sale') AND (retrieve_sq_ft(dataset."facts and features") <> 0)
more>GROUP BY retrieve_bedrooms(dataset."facts and features")
more>ORDER BY retrieve_bedrooms(dataset."facts and features") ASC ;
```

bedrooms	average_price_per_sqft
0	1250
2	715.6112156567557
3	678.5276650040598
4	988.5857436766307
5	908.3493182886695
6	421.55555555555554
7	1126
8	1567
9	1108
11	433

10 tuples

Figure 11