

SELECT \*  
FROM 'billionaires.csv';

|   | rank int64<br>1 - 497 | finalWorth int64<br>5300 - 211000 | category object<br>Finance & In... 16.2%<br>Technology 13.1%<br>16 others 70.7% | personName object<br>Bernard Arn... 0.2%<br>Elon Musk 0.2%<br>473 others 99.6% | age int64<br>30 - 96 | country object<br>United States 40%<br>China 15.4%<br>37 others 44.6% | city object<br>New York 5.1%<br>London 3.6%<br>219 others 91.4% |   |
|---|-----------------------|-----------------------------------|---|--|----------------------|---|---|---|
| 0 | 1                     | 211000                            | Fashion & Retail  | Bernard Arnault &...   | 74                   | France  | Paris   | L |
| 1 | 2                     | 180000                            | Automotive  | Elon Musk  | 51                   | United States   | Austin  | T |
| 2 | 3                     | 114000                            | Technology  | Jeff Bezos   | 59                   | United States   | Medina  | A |
| 3 | 4                     | 107000                            | Technology  | Larry Ellison  | 78                   | United States   | Lanai   | C |
| 4 | 5                     | 106000                            | Finance & Invest...   | Warren Buffett   | 92                   | United States   | Omaha   | E |
| 5 | 6                     | 104000                            | Technology  | Bill Gates   | 67                   | United States   | Medina  | M |
| 6 | 7                     | 94500                             | Media & Entertain...  | Michael Bloomberg  | 81                   | United States   | New York  | E |
| 7 | 8                     | 93000                             | Telecom   | Carlos Slim Helu ...   | 83                   | Mexico  | Mexico City   | T |
| 8 | 9                     | 83400                             | Diversified   | Mukesh Ambani  | 65                   | India   | Mumbai  | D |
| 9 | 10                    | 80700                             | Technology  | Steve Ballmer  | 67                   | United States   | Hunts Point   | N |

475 rows, showing 10 per page

<< < Page 1 of 48 > >>

```
/*  
-- regular expression  
SELECT * FROM 'billionaires.csv'  
WHERE country = 'France'; --case sensitive  
*/  
  
/*  
-- countries starting with the letter F  
SELECT * FROM 'billionaires.csv'  
WHERE country ILIKE 'F%'; -- equivalent to REGEXP '^f' as in the video  
*/  
  
/*  
-- countries ending with the letter A  
SELECT * FROM 'billionaires.csv'  
WHERE country ILIKE '%A'; -- equivalent to REGEXP 'A$' as in the video  
*/  
  
/*  
-- countries starting with the letter F or C  
SELECT * FROM 'billionaires.csv'  
WHERE country ILIKE 'F%' OR country ILIKE 'C%'; -- equivalent to REGEXP '^f|^c' as in the video  
*/  
  
-- countries starting with the letter C then have H or O  
SELECT * FROM 'billionaires.csv'  
WHERE country ILIKE 'Ch%' OR country ILIKE 'Co%'; -- equivalent to REGEXP '^c[ho]' as in the video
```

|   | rank int64<br>15 - 486 | finalWorth int64<br>5400 - 68000 | category object<br>Technology 18.7%<br>Manufacturi... 17.3%<br>11 others 64% | personName object<br>Zhong Shan... 1.3%<br>Zhang Yiming 1.3%<br>73 others 97.3% | age int64<br>39 - 90 | country object<br>China 97.3%<br>Chile 1.3%<br>Colombia 1.3% | city object<br>Shenzhen 13.3%<br>Beijing 12%<br>25 others 74.7% |    |
|---|------------------------|----------------------------------|--|---|----------------------|--|---|----|
| 0 | 15                     | 68000                            | Food & Beverage  | Zhong Shanshan  | 68                   | China  | Hangzhou  | E  |
| 1 | 26                     | 45000                            | Technology   | Zhang Yiming  | 39                   | China  | Beijing   | T  |
| 2 | 34                     | 35300                            | Technology   | Ma Huateng  | 51                   | China  | Shenzhen  | Ir |
| 3 | 38                     | 33400                            | Automotive   | Robin Zeng  | 54                   | China  | Ningde  | E  |
| 4 | 45                     | 30200                            | Technology   | Colin Zheng Huang   | 43                   | China  | Shanghai  | E  |
| 5 | 53                     | 26700                            | Technology   | William Ding  | 51                   | China  | Hangzhou  | C  |
| 6 | 63                     | 23500                            | Technology   | Jack Ma   | 58                   | China  | Hangzhou  | E  |
| 7 | 64                     | 23400                            | Manufacturing  | He Xiangjian & fa...  | 80                   | China  | Foshan  | F  |
| 8 | 65                     | 23100                            | Metals & Mining  | Iris Fontbona & fa...   | 80                   | Chile  | Santiago  | M  |
| 9 | 74                     | 21200                            | Service  | Wang Wei  | 52                   | China  | Shenzhen  | P  |

75 rows, showing 10 per page

<< < Page 1 of 8 > >>

```
-- categorisation by high or low with a threshold - IF
SELECT personName, finalWorth,
IF (finalWorth < 10000, 'low', 'high') AS category -- to make the column name cleaner
FROM 'billionaires.csv';
```

|     | personName object   | finalWorth int64 | category object                     |  |
|-----|---|------------------|-------------------------------------|--|
|     | Bernard Arn... 0.2%<br>Elon Musk ..... 0.2%<br>473 others ..... 99.6% | 5300 - 211000    | low ..... 64.2%<br>high ..... 35.8% |  |
| 470 | Gary Rollins  | 5300             | low                                 |  |
| 471 | Mark Walter   | 5300             | low                                 |  |
| 472 | Ronald Wanek  | 5300             | low                                 |  |
| 473 | Erich Wesjohann ...   | 5300             | low                                 |  |
| 474 | M.A. Yusuff Ali   | 5300             | low                                 |  |

475 rows, showing 10 per page << < Page 48 of 48 > >> [↓](#)

```
-- three categories - CASE
SELECT personName, finalWorth,
CASE
    WHEN finalWorth < 10000 THEN 'low'
    WHEN finalWorth BETWEEN 10000 AND 30000 THEN 'medium'
    ELSE 'high'
END
FROM 'billionaires.csv';
```

|     | personName object   | finalWorth int64 | CASE WHEN ((fina...                                      |  |
|-----|---|------------------|--|--|
|     | Bernard Arn... 0.2%<br>Elon Musk ..... 0.2%<br>473 others ..... 99.6% | 5300 - 211000    | low ..... 64.2%<br>medium ..... 26.5%<br>high ..... 9.3% |  |
| 470 | Gary Rollins  | 5300             | low  |  |
| 471 | Mark Walter   | 5300             | low  |  |
| 472 | Ronald Wanek  | 5300             | low  |  |
| 473 | Erich Wesjohann ...   | 5300             | low  |  |
| 474 | M.A. Yusuff Ali   | 5300             | low  |  |

475 rows, showing 10 per page << < Page 48 of 48 > >> [↓](#)

SUBQUERY: a query inside another query

```
-- Subquery

/*
-- show average net worth
SELECT AVG(finalWorth)
FROM 'billionaires.csv';
*/

SELECT personname, finalWorth,
--nested query
(SELECT AVG(finalWorth) FROM 'billionaires.csv')
FROM 'billionaires.csv';
```

|   | personName object   | finalWorth int64 | (SELECT avg(final... |  |
|---|---|------------------|----------------------|--|
|   | Bernard Arn... 0.2%<br>Elon Musk ..... 0.2%<br>473 others ..... 99.6% | 5300 - 211000    | 14821.8947368421...  |  |
| 0 | Bernard Arnault &...  | 211000           | 14821.89474          |  |
| 1 | Elon Musk   | 180000           | 14821.89474          |  |
| 2 | Jeff Bezos  | 114000           | 14821.89474          |  |
| 3 | Larry Ellison   | 107000           | 14821.89474          |  |
| 4 | Warren Buffett  | 106000           | 14821.89474          |  |
| 5 | Bill Gates  | 104000           | 14821.89474          |  |
| 6 | Michael Bloomberg   | 94500            | 14821.89474          |  |
| 7 | Carlos Slim Helu ...  | 93000            | 14821.89474          |  |
| 8 | Mukesh Ambani   | 83400            | 14821.89474          |  |
| 9 | Steve Ballmer   | 80700            | 14821.89474          |  |

475 rows, showing 10 per page << < Page 1 of 48 > >> [↓](#)

```
-- bring in another csv or table
SELECT * FROM 'geography.csv'
```

|   | ID int64<br>1 - 38<br> | Country_Name obj...  | Language object  | Continent object   |  |
|---|------------------------|--|--|--|--|
|   |                        | Mexico ..... 2.6%<br>India ..... 2.6%<br>36 others ..... 94.7% | English ..... 15.8%<br>Spanish ..... 7.9%<br>25 others ..... 76.3% | Europe ..... 44.7%<br>Asia ..... 31.6%<br>4 others ..... 23.7% |  |
| 0 | 1                      | Mexico   | Spanish  | North America  |  |
| 1 | 2                      | India  | Hindi  | Asia   |  |
| 2 | 3                      | Spain  | Spanish  | Europe   |  |
| 3 | 4                      | China  | Chinese  | Asia   |  |
| 4 | 5                      | Canada   | English  | North America  |  |
| 5 | 6                      | Germany  | German   | Europe   |  |
| 6 | 7                      | Andorra  | Catalan  | Europe   |  |
| 7 | 8                      | Portugal   | Portuguese   | Europe   |  |
| 8 | 9                      | Nepal  | Nepali   | Asia   |  |
| 9 | 10                     | Switzerland  | German   | Europe   |  |

38 rows, showing 10 per page << < Page 1 of 4 > >> [↓](#)

```
-- combine two tables / csv files - joins

/*
--inner join
SELECT * FROM 'billionaires.csv'
INNER JOIN 'geography.csv'
ON billionaires.country = geography.Country_Name;
*/

/*
-- bit cleaner
--inner join
SELECT country, Continent, Language FROM 'billionaires.csv'
INNER JOIN 'geography.csv'
ON billionaires.country = geography.Country_Name;
*/

/*
--left join
SELECT Distinct(country), Continent, Language FROM 'billionaires.csv' -- remove duplicates
LEFT JOIN 'geography.csv'
ON billionaires.country = geography.Country_Name;
*/

/*
--right join
SELECT Distinct(country), Continent, Language FROM 'billionaires.csv' -- remove duplicates
RIGHT JOIN 'geography.csv'
ON billionaires.country = geography.Country_Name;
*/

--right join - observe and test more
SELECT Distinct(country), Country_Name, Continent, Language FROM 'billionaires.csv' -- remove duplicates
RIGHT JOIN 'geography.csv'
ON billionaires.country = geography.Country_Name;
```

|   | country object   | Country_Name obj...  | Continent object   | Language object  |  |
|---|--|--|--|--|--|
|   | Mexico ..... 2.6%<br>34 others ..... 89.5%<br>Missing ..... 7.9% | Mexico ..... 2.6%<br>India ..... 2.6%<br>36 others ..... 94.7% | Europe ..... 44.7%<br>Asia ..... 31.6%<br>4 others ..... 23.7% | English ..... 15.8%<br>Spanish ..... 7.9%<br>25 others ..... 76.3% |  |
| 0 | Mexico   | Mexico   | North America  | Spanish  |  |
| 1 | India  | India  | Asia   | Hindi  |  |
| 2 | Spain  | Spain  | Europe   | Spanish  |  |
| 3 | China  | China  | Asia   | Chinese  |  |
| 4 | Canada   | Canada   | North America  | English  |  |
| 5 | Germany  | Germany  | Europe   | German   |  |
| 6 | Switzerland  | Switzerland  | Europe   | German   |  |
| 7 | Belgium  | Belgium  | Europe   | Belgian  |  |
| 8 | Austria  | Austria  | Europe   | Austrian   |  |
| 9 | Japan  | Japan  | Asia   | Japanese   |  |

38 rows, showing 10 per page << < Page 1 of 4 > >> [↓](#)

