

N(movie)	240
N(person)	5000
N(movie_cast)	3179
V(year, movie)	54
Min(year, movie)	1927
Max(year, movie)	2014
V(cast_order, movie_cast)	65

1.

SELECT * FROM movie CROSS JOIN person

Estimativa:

Sendo um cross join, tudo será unido com tudo, portanto serão 240 registros de movies * 5000 registros de person = 1.200.000 linhas.

Resultado real:

1.200.000 rows.

2.

**SELECT * FROM movie NATURAL JOIN movie_cast
WHERE cast_order = 1 AND release_year > 2000**

Estimativa:

Temos 65 valores distintos pra cast order, portanto a chance da tupla encontrada ter cast_order = 1 é de 1/65. Além disso, para os anos, temos $(2000 - 1927 + 1) / (2014 - 1927 + 1) = 1 - (74 / 88) = 14/88 = 7/44$. Com isso, são $14/88 * 1/65 * 3179 = 7,78$

Resultado real:

87 rows

Resultado muito maior que a estimativa, visto que na base de dados a distribuição de filmes por ano é dispersa, coisa que a estimativa não trata.

3.

```
SELECT * FROM movie JOIN movie_cast USING (movie_id)  
JOIN person USING (person_id)  
WHERE cast_order = 1 AND (release_year = 1990 OR release_year = 2000)
```

Estimativa pra questao 3:

$$2/54 * 1/65 * 3179 = 1.81$$

Resultado real:

18 rows

Resultado novamente muito maior que a estimativa, o que significa que um pico de disparidade no lançamento de filmes se encontra nesse intervalo, de acordo com nosso BD.

4.

```
SELECT title FROM movie UNION ALL  
SELECT character_name FROM movie_cast
```

Estimativa pra questao 4:

Basta juntar a quantidade de filmes com a quantidade de linhas no movie_cast.

$$\text{Então: } 3179 + 240 = 3419$$

Resultado real:

3419 rows

O resultado equivale a soma das tuplas de cada tabela.

5.

```
SELECT title FROM movie EXCEPT  
SELECT character_name FROM movie_cast
```

Estimativa pra questao 5:

Aqui, devemos seguir a lógica de pegar a quantidade de filmes também. Como não temos como supor a quantidade de casos que o character_name de algum personagem é igual ao título de algum filme, resta a nós tomar algum viés, e faz sentido manter-nos com os 240 filmes como quantidade final, sabendo que a probabilidade de ser próximo do resultado certo será alta.

Resultado = 227 rows

Na estimativa assumimos que nenhum filme teria como título o nome de um personagem. Apesar de sabermos que isso é falso, é o mais próximo que

conseguimos chegar estimando e, de fato, perto chegamos. A busca no banco apenas mostra que em alguns casos isso de fato ocorre.