

SQL Interview

1. SELECT

 *

 FROM

 movie

;

2. SELECT

 title

 FROM

 movie

 ORDER BY

 runtime ASC

;

3. SELECT

 title

 FROM

 movie

 ORDER BY

 revenue ASC

 LIMIT 1

;

4. SELECT

 title

 FROM

 movie

 ORDER BY

 revenue ASC

;

5. SELECT

 title, gender_id, person_id, character_name, cast_order

 FROM

 movie

 LEFT JOIN

 movie_cast

 ON

 movie.movie_id = movie_cast.movie_id

;

6. SELECT

 country_name, COUNT(movie_id) AS no_of_movies

 FROM

```

        country
LEFT JOIN production_country
    Country.country_id = production_country.country_id
GROUP BY
    country_id
ORDER BY
    no_of_movies
LIMIT 1
;

```

7. SELECT

```

    *
FROM
    genre
;

```

8. SELECT

```

    language_id, COUNT(movie_id) AS no_of_movies
FROM
    movie_language
;

```

9. SELECT

```

    title,
    COUNT(DISTINCT person_id.movie_cast) AS no_of_cast,
    COUNT(DISTINCT person_id.movie_crew) AS no_of_crew
FROM
    ((movie
    INNER JOIN movie_cast ON movie_id.movie = movie_id.movie_cast)
    INNER JOIN movie_crew ON movie_id.movie = movie_id.movie_crew)
;

```

10. SELECT

```

    title
FROM
    movie
ORDER BY
    popularity DESC
LIMIT 10
;

```

11. SELECT

```

    title, revenue
FROM
    movie
LIMIT 2,1
;

```

```
12. SELECT
    title
FROM
    movie
WHERE
    status="rumoured"
;
```

```
13. SELECT
    title.movie
FROM
    ((movie
    INNER JOIN production_country ON movie.movie_id =
    production_country.movie_id)
    INNER JOIN country ON production_country.country_id = country.country_id)
ORDER BY
    revenue.movie DESC
LIMIT 1
;
```

```
14. SELECT
    movie_id.movie_company,company_name.production_company
FROM
    movie_company,production_company
WHERE
    movie_company.movie_id = production_company.movie_id
;
```

```
15. SELECT
    title
FROM
    movie
ORDER BY
    budget DESC
;
```

Statistics Questions

1. D
2. C
3. C
4. D
5. C
6. B
7. A
8. A
9. B
10. A

Machine Learning Questions

1. The RSS is just the absolute amount of explained variation, the R squared is the (RSS/SST), i.e. the absolute amount of variation as a proportion of total variation.
2. $TSS = ESS + RSS$, where TSS is Total Sum of Squares, ESS is Explained Sum of Squares and RSS is Residual Sum of Squares. The aim of Regression Analysis is explain the variation of dependent variable Y.
3. Regularization is used to overcome the problem of overfitting.
4. Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen.
5. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.
6. Ensemble methods are techniques that create multiple models and then combine them to produce improved results. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.
7. Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance. Moreover, in bagging the data flows in a parallel manner whereas in boosting model data flows in series from one model to another.
- 8.
9. Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation.
10. Hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process.
11. Having a larger learning rate could lead to skipping of the ultimate minima point hence we could get the values at the lowest possible point.
- 12.