

Plan de Preparación – SQL & Pandas para Entrevista Data Engineer

Duración: 3 meses (12 semanas)

Frecuencia: 2 días por semana

Perfil: Ingeniero / Backend con experiencia práctica

Objetivo: Dominar SQL de entrevistas y análisis de datos con Pandas

Semana 1

Día 1: SQL Básico – SELECT & WHERE

- Employees Earning More Than Their Managers
- Second Highest Salary
- Combine Two Tables
- Duplicate Emails
- Customers Who Never Order
- Big Countries
- Recyclable and Low Fat Products
- Invalid Tweets
- Article Views I
- Find Customer Referee

Día 2: SQL Agregaciones

- Average Time of Process per Machine
- Project Employees I
- Game Play Analysis I
- Immediate Food Delivery I
- Number of Unique Subjects Taught by Each Teacher
- Queries Quality and Percentage
- Monthly Transactions I
- Customer Placing the Largest Number of Orders
- Classes More Than 5 Students
- User Activity for the Past 30 Days I

Semana 2

Día 3: JOINs

- Rising Temperature
- Employees Earning More Than Their Managers
- Sales Person
- Triangle Judgement

- Customer Who Visited but Did Not Make Any Transactions
- Replace Employee ID With The Unique Identifier
- Managers with at Least 5 Direct Reports
- Investments in 2016
- Market Analysis I
- Market Analysis II

Día 4: Subqueries & CTEs

- Department Highest Salary
- Nth Highest Salary
- Rank Scores
- Consecutive Numbers
- Trips and Users
- Exchange Seats
- Median Employee Salary
- Department Top Three Salaries
- Human Traffic of Stadium
- Active Businesses

Semana 3

Día 5: Window Functions

- Rank Scores
- Department Top Three Salaries
- Running Total for Different Genders
- Last Person to Fit in the Bus
- Cumulative Salary
- Average Selling Price
- Game Play Analysis IV
- User Purchase Platform
- Monthly Transactions II
- Restaurant Growth

Día 6: Fechas & Time Series

- Game Play Analysis I
- Game Play Analysis II
- Game Play Analysis III
- Reported Posts
- User Activity for the Past 30 Days I
- User Activity for the Past 30 Days II
- Ad-Free Sessions
- Monthly Transactions I
- Monthly Transactions II
- Reported Posts II

Semana 4

Día 7: Data Cleaning SQL

- Fix Names in a Table
- Delete Duplicate Emails
- Remove Duplicate Emails
- Reformat Department Table
- Find Total Time Spent by Each Employee
- Group Sold Products By The Date
- Unpopular Books
- Customers With Strictly Increasing Purchases
- Products Price for Each Store
- Warehouse Manager

Día 8: Pandas Básico

- DataFrame Creation and Indexing
- Filtering Rows with Conditions
- Handling Missing Values
- Rename Columns
- Sorting Data
- Value Counts
- Apply Functions
- String Operations
- Date Parsing
- Export DataFrames

Semana 5

Día 9: Pandas Aggregations

- GroupBy Mean and Sum
- GroupBy Multiple Columns
- Aggregations with Named Agg
- Pivot Tables
- Cumulative Sum
- Rolling Average
- Ranking Data
- Percentage Change
- Outlier Detection
- Custom Aggregations

Día 10: Pandas Joins

- Merge Two DataFrames
- Left / Right / Inner Join

- Join on Multiple Keys
- Handling Duplicates After Merge
- Concat DataFrames
- Union Operations
- Compare DataFrames
- Anti Join Logic
- Cross Join
- Join with Time Series

Semana 6

Día 11: SQL + Pandas (ETL)

- Load SQL Data into Pandas
- Clean Data and Normalize
- Build Fact Table Logic
- Build Dimension Tables
- Detect Duplicates
- Incremental Loads
- Validate Data Quality
- Generate Metrics Table
- Export to Data Warehouse
- End-to-End Mini ETL

Día 12: Mock Interview SQL

- Design Query for Daily Active Users
- Top N Products per Category
- Detect Fraudulent Transactions
- User Retention Query
- Funnel Analysis
- Sessionization Query
- Revenue Growth MoM
- Data Quality Checks
- Slowly Changing Dimensions
- Explain Query Optimization

Semana 7

Día 13: SQL Básico – SELECT & WHERE

- Employees Earning More Than Their Managers
- Second Highest Salary
- Combine Two Tables
- Duplicate Emails
- Customers Who Never Order
- Big Countries

- Recyclable and Low Fat Products
- Invalid Tweets
- Article Views I
- Find Customer Referee

Día 14: SQL Agregaciones

- Average Time of Process per Machine
- Project Employees I
- Game Play Analysis I
- Immediate Food Delivery I
- Number of Unique Subjects Taught by Each Teacher
- Queries Quality and Percentage
- Monthly Transactions I
- Customer Placing the Largest Number of Orders
- Classes More Than 5 Students
- User Activity for the Past 30 Days I

Semana 8

Día 15: JOINs

- Rising Temperature
- Employees Earning More Than Their Managers
- Sales Person
- Triangle Judgement
- Customer Who Visited but Did Not Make Any Transactions
- Replace Employee ID With The Unique Identifier
- Managers with at Least 5 Direct Reports
- Investments in 2016
- Market Analysis I
- Market Analysis II

Día 16: Subqueries & CTEs

- Department Highest Salary
- Nth Highest Salary
- Rank Scores
- Consecutive Numbers
- Trips and Users
- Exchange Seats
- Median Employee Salary
- Department Top Three Salaries
- Human Traffic of Stadium
- Active Businesses

Semana 9

Día 17: Window Functions

- Rank Scores
- Department Top Three Salaries
- Running Total for Different Genders
- Last Person to Fit in the Bus
- Cumulative Salary
- Average Selling Price
- Game Play Analysis IV
- User Purchase Platform
- Monthly Transactions II
- Restaurant Growth

Día 18: Fechas & Time Series

- Game Play Analysis I
- Game Play Analysis II
- Game Play Analysis III
- Reported Posts
- User Activity for the Past 30 Days I
- User Activity for the Past 30 Days II
- Ad-Free Sessions
- Monthly Transactions I
- Monthly Transactions II
- Reported Posts II

Semana 10

Día 19: Data Cleaning SQL

- Fix Names in a Table
- Delete Duplicate Emails
- Remove Duplicate Emails
- Reformat Department Table
- Find Total Time Spent by Each Employee
- Group Sold Products By The Date
- Unpopular Books
- Customers With Strictly Increasing Purchases
- Products Price for Each Store
- Warehouse Manager

Día 20: Pandas Básico

- DataFrame Creation and Indexing
- Filtering Rows with Conditions
- Handling Missing Values
- Rename Columns

- Sorting Data
- Value Counts
- Apply Functions
- String Operations
- Date Parsing
- Export DataFrames

Semana 11

Día 21: Pandas Aggregations

- GroupBy Mean and Sum
- GroupBy Multiple Columns
- Aggregations with Named Agg
- Pivot Tables
- Cumulative Sum
- Rolling Average
- Ranking Data
- Percentage Change
- Outlier Detection
- Custom Aggregations

Día 22: Pandas Joins

- Merge Two DataFrames
- Left / Right / Inner Join
- Join on Multiple Keys
- Handling Duplicates After Merge
- Concat DataFrames
- Union Operations
- Compare DataFrames
- Anti Join Logic
- Cross Join
- Join with Time Series

Semana 12

Día 23: SQL + Pandas (ETL)

- Load SQL Data into Pandas
- Clean Data and Normalize
- Build Fact Table Logic
- Build Dimension Tables
- Detect Duplicates
- Incremental Loads
- Validate Data Quality
- Generate Metrics Table

- Export to Data Warehouse
- End-to-End Mini ETL

Día 24: Mock Interview SQL

- Design Query for Daily Active Users
- Top N Products per Category
- Detect Fraudulent Transactions
- User Retention Query
- Funnel Analysis
- Sessionization Query
- Revenue Growth MoM
- Data Quality Checks
- Slowly Changing Dimensions
- Explain Query Optimization