A blue and white logo

AI-generated content may be incorrect.

SQL ASSESSMENT WORKBOOK 3

V1: **CUSTOMER RISK SEGMENTATION**

**EduFin SQL Skill Assessment Workbook:  
*A Beginner-Friendly Guide to Customer Segmentation, Scoring Models, and Cohort-Based Analytical Patterns***

**Program: Skill AI Path – Data Analyst Pretraining Track  
Module: EduFin Risk Analytics Simulation – Session 3  
Assessment Type: Skill Validation – Segmentation Logic + CASE Scoring + Statistical Cohort Analysis  
Prepared For: EduFin Data Analyst Cohort**

**Objective**

To assess and validate your ability to perform advanced customer analytics using SQL, including multi-dimensional segmentation, CASE-based scoring systems, income distribution analysis, and cohort behavior tracking—preparing you for real-world customer profiling and insight generation in data-driven financial environments.

# SQL Skill Check Assessment – Pretraining WorkbooK

**SKILL CHECK ASSESSMENT**

This skill check is designed to assess your ability to handle advanced multi-dimensional analysis, scoring logic, cohort analytics, and statistical segmentation using SQL.

## Part A: Query Writing.

**Question 1 (10 points):**  
Create a multi-dimensional customer segmentation showing employment type, income bracket (create 4 brackets), and customer segment, with count, average income, and loan penetration rate for each combination.

**Answer:**

**Question 2 (10 points):**  
Build a customer risk scoring system using CASE statements that considers annual income (40% weight), employment type (30% weight), customer segment (20% weight), and city tier (10% weight). Show customers with their total scores and risk classifications.

**Answer:**

**Question 3 (10 points):**  
Perform a cross-tabulation analysis showing customer segments as rows and loan status as columns, displaying counts and percentages for each combination.

**Answer:**

**Question 4 (10 points):**  
Create a cohort analysis showing customers grouped by registration quarter, with their subsequent loan behavior including average loan amounts, default rates, and portfolio growth.

**Answer:**

**Question 5 (10 points):**  
Write a query using statistical functions (PERCENTILE\_CONT, STDEV) to analyze income distribution within each employment type and identify which employment types have the most income variance.

**Answer:**

**Question 6 (10 points):**  
Build a customer ranking system that ranks customers within their employment type based on total portfolio value adjusted for default history, showing top 5 customers per employment type.

**Answer:**

## Part B: Multiple Choice Questions

**Total: 40 points (10 questions × 4 points)**

**Instructions:** Choose the correct option. Mark only one option as your answer for each question.

**Question 7:** Which function calculates the median value of a dataset?

* A) AVG()
* B) PERCENTILE\_CONT(0.5)
* C) MEDIAN()
* D) MIDDLE()

**Answer:** \_\_\_

**Question 8:** In multi-dimensional GROUP BY, what determines the granularity of results?

* A) ORDER BY clause
* B) Number of columns in GROUP BY
* C) HAVING clause
* D) SELECT clause

**Answer:** \_\_\_

**Question 9:** What does NTILE(10) do?

* A) Returns top 10 records
* B) Divides data into 10 equal groups
* C) Calculates 10th percentile
* D) Creates 10 random samples

**Answer:** \_\_\_

**Question 10:** Which is correct for conditional aggregation?

* A) COUNT(CASE WHEN condition THEN 1 END)
* B) COUNT(WHERE condition)
* C) COUNT(IF condition THEN 1)
* D) COUNT(condition = TRUE)

**Answer:** \_\_\_

**Question 11:** What does STDEV() measure?

* A) Standard error
* B) Standard deviation (variability)
* C) Statistical significance
* D) Sample size

**Answer:** \_\_\_

**Question 12:** In customer segmentation, why use HAVING COUNT(\*) >= 10?

* A) Improves performance
* B) Ensures statistical significance
* C) Required by SQL
* D) Prevents NULL values

**Answer:** \_\_\_

**Question 13:** Which creates a proper customer risk score with weighted factors?

* A) SUM(factor1 + factor2 + factor3)
* B) (factor1 \* 0.4) + (factor2 \* 0.3) + (factor3 \* 0.3)
* C) AVG(factor1, factor2, factor3)
* D) MAX(factor1, factor2, factor3)

**Answer:** \_\_\_

**Question 14:** What's the difference between ROW\_NUMBER() and RANK()?

* A) No difference
* B) ROW\_NUMBER() gives unique ranks, RANK() allows ties
* C) RANK() is faster
* D) ROW\_NUMBER() handles NULLs better

**Answer:** \_\_\_

**Question 15:** Why use COALESCE() in aggregations?

* A) Improves performance
* B) Handles NULL values in calculations
* C) Required for GROUP BY
* D) Prevents duplicate counting

**Answer:** \_\_\_

**Question 16:** Which pattern creates effective customer cohorts?

* A) GROUP BY customer\_id
* B) GROUP BY registration\_date
* C) GROUP BY YEAR(registration\_date), QUARTER(registration\_date)
* D) GROUP BY customer\_segment

**Answer:** \_\_\_

## Passing Criteria

* **Minimum Score:** 80 out of 100
* **Query Writing Section:** At least 48/60
* **MCQ Section:** At least 32/40
* **Unlimited Attempts:** Yes
* **Time Limit:** 150 minutes per attempt

## Skills You Will Validate

By completing this session, you demonstrate:

* Advanced use of multi-column GROUP BY
* Comfort with statistical and distributional functions
* Ability to segment and score customers logically
* Ranking and cross-tab analytics using window functions
* Cohort trend identification and growth analytics
* Risk-adjusted value computation and prioritization

## Study Recommendations Before Attempting Again

1. Review CASE scoring techniques
2. Practice cohort-based grouping using quarter and month
3. Explore window functions like RANK(), ROW\_NUMBER(), NTILE()
4. Use STDEV and PERCENTILE\_CONT in practice queries
5. Apply customer-level logic across multiple attributes

© 2025 Skill AI Path. All Rights Reserved.  
Licensed for educational and internal business use.  
Attribution required when sharing. Commercial redistribution prohibited.

Contact Us - [tech@skillaipath.com](mailto:tech@skillaipath.com) | [skillaipath.com](http://skillaipath.com/)