Marco 2 do Projeto: Classificação em Base ao Risco Relativo

Maior Risco de Crédito:

```
CREATE OR REPLACE TABLE risco-relativo.credito.mau AS
WITH dados AS (
  SELECT
    age,
    more_90_days_overdue,
    using_lines_not_secured_personal_assets,
    total_loan,
    --clean_loan_type,
    number_dependents_median,
    last_month_salary_median,
    default_flag,
    NTILE(4) OVER (ORDER BY age) AS quartil_idade,
    NTILE(4) OVER (ORDER BY more_90_days_overdue) AS quartil_days,
    NTILE(4) OVER (ORDER BY using_lines_not_secured_personal_assets) AS
quartil_ativo,
    NTILE(4) OVER (ORDER BY total_loan) AS quartil_emprestimos,
    --CASE WHEN clean_loan_type = 'Real Estate' THEN 1 ELSE 0 END AS tipo_credito,
    NTILE(4) OVER (ORDER BY number_dependents_median) AS quartil_dependente,
    NTILE(4) OVER (ORDER BY last_month_salary_median) AS quartil_salario
  FROM
    `risco-relativo.credito.full_join`
),

riscos AS (
  SELECT
    'Idade' AS variavel,
    quartil_idade AS quartil,,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join` ) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Idade' ORDER BY AVG(CAST(default_flag AS
FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(age) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Dias de Atraso' AS variavel,
    quartil_days AS quartil,
```

```
    COUNT(more_90_days_overdue) AS quantidade,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Dias de Atraso' ORDER BY AVG(CAST(default_flag
AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(more_90_days_overdue) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Uso Limite de Credito' AS variavel,
    quartil_ativo AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Uso Limite de Credito' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(using_lines_not_secured_personal_assets) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Total de Empréstimos' AS variavel,
    quartil_emprestimos AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Total de Empréstimos' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`)DESC) AS rn,
    AVG(total_loan) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
```

```
    'Dependente' AS variavel,
    quartil_dependente AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Total de Empréstimos' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`)DESC) AS rn,
    AVG(number_dependents_median) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Salario' AS variavel,
    quartil_salario AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Salario' ORDER BY AVG(CAST(default_flag AS
FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(last_month_salary_median) AS valor
  FROM
    dados
  GROUP BY
    quartil
),

maiores_riscos AS (
  SELECT
    variavel,
    quartil,
    valor,
    incidencia,
    risco_relativo
  FROM
    riscos
  WHERE
    rn = 1
)

SELECT * FROM maiores_riscos ORDER BY risco_relativo DESC;
```

Menor Risco de Crédito:

```
CREATE OR REPLACE TABLE risco-relativo.credito.bons AS
WITH dados AS (
  SELECT
    age,
    more_90_days_overdue,
    using_lines_not_secured_personal_assets,
    total_loan,
    --clean_loan_type,
    number_dependents_median,
    last_month_salary_median,
    default_flag,
    NTILE(4) OVER (ORDER BY age) AS quartil_idade,
    NTILE(4) OVER (ORDER BY more_90_days_overdue) AS quartil_days,
    NTILE(4) OVER (ORDER BY using_lines_not_secured_personal_assets) AS
quartil_ativo,
    NTILE(4) OVER (ORDER BY total_loan) AS quartil_emprestimos,
    --CASE WHEN clean_loan_type = 'Real Estate' THEN 1 ELSE 0 END AS tipo_credito,
    NTILE(4) OVER (ORDER BY number_dependents_median) AS quartil_dependente,
    NTILE(4) OVER (ORDER BY last_month_salary_median) AS quartil_salario
  FROM
    `risco-relativo.credito.full_join`
),

riscos AS (
  SELECT
    'Idade' AS variavel,
    quartil_idade AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join` ) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Idade' ORDER BY AVG(CAST(default_flag AS
FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(age) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Dias de Atraso' AS variavel,
    quartil_days AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
```

```
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Dias de Atraso' ORDER BY AVG(CAST(default_flag
AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(more_90_days_overdue) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Uso Limite de Credito' AS variavel,
    quartil_ativo AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Uso Limite de Credito' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(using_lines_not_secured_personal_assets) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Total de Empréstimos' AS variavel,
    quartil_emprestimos AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Total de Empréstimos' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`)DESC) AS rn,
    AVG(total_loan) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Dependente' AS variavel,
    quartil_dependente AS quartil,
```

```
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Total de Empréstimos' ORDER BY
AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`)DESC) AS rn,
    AVG(number_dependents_median) AS valor
  FROM
    dados
  GROUP BY
    quartil
  UNION ALL
  SELECT
    'Salario' AS variavel,
    quartil_salario AS quartil,
    AVG(CAST(default_flag AS FLOAT64)) AS incidencia,
    AVG(CAST(default_flag AS FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64))
FROM `risco-relativo.credito.full_join`) AS risco_relativo,
    ROW_NUMBER() OVER (PARTITION BY 'Salario' ORDER BY AVG(CAST(default_flag AS
FLOAT64)) / (SELECT AVG(CAST(default_flag AS FLOAT64)) FROM
`risco-relativo.credito.full_join`) DESC) AS rn,
    AVG(last_month_salary_median) AS valor
  FROM
    dados
  GROUP BY
    quartil
),

menores_riscos AS (
  SELECT
    variavel,
    quartil,
    valor,
    incidencia,
    risco_relativo
  FROM
    riscos
  WHERE
    rn = 4
)

SELECT * FROM menores_riscos oRDER BY risco_relativo DESC;
```

Query Consultas:

```sql
—segmentaçao em base a risco relativo
WITH dados AS (
  SELECT
    user_id,
    age,
    more_90_days_overdue,
    number_times_delayed_payment_loan_30_59_days,
    number_times_delayed_payment_loan_60_89_days,
    total_loan,
    last_month_salary_median,
    number_dependents_median,
    using_lines_not_secured_personal_assets,
    NTILE(4) OVER (ORDER BY age) AS quartil_idade,
    NTILE(4) OVER (ORDER BY more_90_days_overdue) AS quartil_90days,
    NTILE(4) OVER (ORDER BY number_times_delayed_payment_loan_30_59_days) AS
quartil_30days,
    NTILE(4) OVER (ORDER BY number_times_delayed_payment_loan_60_89_days) AS
quartil_60days,
    NTILE(4) OVER (ORDER BY total_loan) AS quartil_empretismo,
    NTILE(4) OVER (ORDER BY last_month_salary_median) AS quartil_salario,
    NTILE(4) OVER (ORDER BY number_dependents_median) AS quartil_dependente,
    NTILE(4) OVER (ORDER BY using_lines_not_secured_personal_assets) AS
quartil_ativo
  FROM
    `risco-relativo.credito.full_join`
),
riscos AS (
  SELECT
    user_id,
    quartil_idade,
    more_90_days_overdue,
    number_times_delayed_payment_loan_30_59_days,
    number_times_delayed_payment_loan_60_89_days,
    quartil_empretismo,
    quartil_salario,
    quartil_dependente,
    quartil_ativo,
    CASE WHEN quartil_idade = 1 THEN 1 ELSE 0 END AS idade_risco,
    --CASE WHEN more_90_days_overdue > 1 THEN 1 ELSE 0 END AS dias90_risco,
    --CASE WHEN number_times_delayed_payment_loan_30_59_days > 1 THEN 1 ELSE 0 END
AS dias30_risco,
    --CASE WHEN number_times_delayed_payment_loan_60_89_days > 1 THEN 1 ELSE 0 END
AS dias60_risco,
```

```sql
    CASE WHEN quartil_90days > 4 THEN 1 ELSE 0 END AS dias90_risco,
    CASE WHEN quartil_30days > 4 THEN 1 ELSE 0 END AS dias30_risco,
    CASE WHEN quartil_60days > 4 THEN 1 ELSE 0 END AS dias60_risco,
    CASE WHEN quartil_empretismo = 1 THEN 1 ELSE 0 END AS empretismo_risco,
    CASE WHEN quartil_salario = 1 THEN 1 ELSE 0 END AS salario_risco,
    CASE WHEN quartil_dependente = 4 THEN 1 ELSE 0 END AS dependente_risco,
    CASE WHEN quartil_ativo = 4 THEN 1 ELSE 0 END AS uso_limite_risco
  FROM
    dados
),
pontuacao AS (
  SELECT
    user_id,
    idade_risco + dias90_risco + dias30_risco + dias60_risco + empretismo_risco +
salario_risco + dependente_risco + uso_limite_risco AS pontuacao
  FROM
    riscos
),

classificacao AS (
  SELECT
    user_id,
    CASE WHEN pontuacao.pontuacao >= 3 THEN 1 ELSE 0 END AS classificacao
  FROM
    pontuacao
)
SELECT
  r.user_id,
  r.age,
  r.more_90_days_overdue,
  r.number_times_delayed_payment_loan_30_59_days,
  r.number_times_delayed_payment_loan_60_89_days,
  r.total_loan,
  r.last_month_salary_median,
  r.number_dependents_median,
  r.using_lines_not_secured_personal_assets,
  p.pontuacao,
  c.classificacao
FROM
  dados r
JOIN
  pontuacao p
ON
  r.user_id = p.user_id
```

```sql
JOIN
  classificacao c
ON
  r.user_id = c.user_id;
--identificar
SELECT COUNT(classificacao)
from risco-relativo.credito.pontuacao_classificacao
WHERE classificacao = 1
```

Matriz de Confusão:

```
--Avaliação e Ajuste: Após classificar os indivíduos com base no risco
relativo(bons, mau), você pode avaliar o desempenho do modelo usando uma matriz de
confusão, por exemplo, para verificar a precisão das classificações e ajustar o
limiar, se necessário, para melhorar o desempenho do modelo.

WITH dados AS (
  SELECT
    age,
    more_90_days_overdue,
    --number_times_delayed_payment_loan_30_59_days,
    --number_times_delayed_payment_loan_60_89_days,
    using_lines_not_secured_personal_assets,
    --last_month_salary_median,
    --clean_loan_type,
    --total_loan,
    default_flag,
    NTILE(4) OVER (ORDER BY age) AS quartil_idade,
    NTILE(4) OVER (ORDER BY more_90_days_overdue) AS quartil_days,
    NTILE(4) OVER (ORDER BY using_lines_not_secured_personal_assets) AS
quartil_ativo,
    --NTILE(4) OVER (ORDER BY total_loan) AS quartil_emprestismo,
    --NTILE(4) OVER (ORDER BY last_month_salary_median) AS quartil_salario,
    --CASE WHEN clean_loan_type = 'Real Estate' THEN 1 ELSE 0 END AS tipo_credito
  FROM
     `risco-relativo.credito.full_join`
),

riscos AS (
  SELECT
    'Idade' AS variavel,
    quartil_idade AS quartil,
    CASE WHEN quartil_idade = 1 THEN 1 ELSE 0 END AS idade_risco,
    CASE WHEN quartil_days = 4 THEN 1 ELSE 0 END AS dias_atraso_risco,
    CASE WHEN quartil_ativo = 4 THEN 1 ELSE 0 END AS uso_limite_risco,
    --CASE WHEN quartil_emprestismo = 1 THEN 1 ELSE 0 END AS emprestismo_risco,
    --CASE WHEN quartil_salario = 1 THEN 1 ELSE 0 END AS salario_risco,
    --CASE WHEN tipo_credito = 0 THEN 1 ELSE 0 END AS tipo_credito_risco
  FROM
    dados
),

pontuacao AS (
  SELECT
```

```
    *,
    idade_risco + dias_atraso_risco + uso_limite_risco AS pontuacao --+
salario_risco + emprestismo_risco + tipo_credito_risco AS pontuacao
  FROM
    riscos
),

classificacao AS (
  SELECT
    *,
    CASE WHEN pontuacao.pontuacao >= 3 THEN 'Mau Pagador' ELSE 'Bom Pagador' END AS
classificacao
  FROM
    pontuacao
),


matriz_confusao AS (
  SELECT
    SUM(CASE WHEN default_flag = 1 AND classificacao.classificacao = 'Mau Pagador'
THEN 1 ELSE 0 END) AS verdadeiros_positivos,
    SUM(CASE WHEN default_flag = 0 AND classificacao.classificacao = 'Bom Pagador'
THEN 1 ELSE 0 END) AS verdadeiros_negativos,
    SUM(CASE WHEN default_flag = 0 AND classificacao.classificacao = 'Mau Pagador'
THEN 1 ELSE 0 END) AS falsos_positivos,
    SUM(CASE WHEN default_flag = 1 AND classificacao.classificacao = 'Bom Pagador'
THEN 1 ELSE 0 END) AS falsos_negativos
  FROM
    classificacao,
    `risco-relativo.credito.full_join`
),

-- Calcular as métricas de avaliação
metricas AS (
  SELECT
    SUM(verdadeiros_positivos) AS tp,
    SUM(verdadeiros_negativos) AS tn,
    SUM(falsos_positivos) AS fp,
    SUM(falsos_negativos) AS fn
  FROM
    matriz_confusao
),
```

```
--f1 = uma métrica que combina precisão e revocação (recall) em um único valor, e um valor de 0.02 indica que o modelo não está muito equilibrado entre essas métricas(+ alto modelo melhor).
precision_recall AS (
  SELECT
    2 * (tp / (tp + fp) * tp / (tp + fn)) / (tp / (tp + fp) + tp / (tp + fn)) AS f1_score
  FROM
    metricas
),

--O recall indica que o modelo está identificando corretamente dos casos positivos(tem q ser + proximo a 1). A taxa de falsos positivos significa que o modelo está classificando incorretamente dos casos negativos como positivos(tem q ser + proximo do 0).
roc_curve AS (
  SELECT
    SUM(verdadeiros_positivos) / (SUM(verdadeiros_positivos) + SUM(falsos_negativos)) AS recall,
    SUM(falsos_positivos) / (SUM(falsos_positivos) + SUM(verdadeiros_negativos)) AS taxa_falsos_positivos
  FROM
    matriz_confusao
)

SELECT
  matriz_confusao.verdadeiros_positivos,
  matriz_confusao.verdadeiros_negativos,
  matriz_confusao.falsos_positivos,
  matriz_confusao.falsos_negativos,
  precision_recall.f1_score,
  roc_curve.recall,
  roc_curve.taxa_falsos_positivos
FROM matriz_confusao, metricas,precision_recall, roc_curve;

--Esses resultados sugerem que o seu modelo está prevendo corretamente metade dos casos positivos (verdadeiros positivos) e metade dos casos negativos (verdadeiros negativos). No entanto, a taxa de falsos positivos é alta, o que indica que o modelo está classificando erroneamente muitos casos negativos como positivos. O baixo F1-score também sugere que o modelo não está performando bem na precisão e revocação das previsões. Isso pode indicar a necessidade de ajustes no modelo ou na estratégia de tratamento dos dados.
```