

# Especificación de Requerimientos de Software (SRS)

## Motor de Recomendación y Dashboard de Campañas Bancarias

Versión 1.5

29 de agosto de 2025

## 1. Introducción

### 1.1. Propósito

Este documento define los requerimientos funcionales y no funcionales para el desarrollo de una aplicación web orientada a:

- La exploración y análisis de datos de campañas de marketing bancario.
- La recomendación de productos financieros (créditos, hipotecas, depósitos a plazo) a clientes en función de su perfil.

### 1.2. Alcance

El sistema se desarrollará en dos fases iterativas:

- **Fase 1:** Dashboard de consultas, reportes y filtros.
- **Fase 2:** Motor de recomendaciones con soporte estadístico para decisiones gerenciales.

**Fuente de datos inicial:** Se provee una base de datos en formato CSV/Excel con un total de 45,211 registros correspondientes a campañas de marketing directo de una entidad bancaria portuguesa. Esta base estará disponible para carga inicial del sistema y podrá ser descargada desde el siguiente enlace:

- **Dataset UCI Bank Marketing:** <https://archive.ics.uci.edu/dataset/222/bank+marketing>

Además, el sitio oficial provee la descripción detallada de los campos, su significado, representación y tipos de valores válidos.

### 1.3. Audiencia prevista

- **Directivos y gerencia:** interesados en KPIs y reportes ejecutivos.
- **Analistas de datos:** usuarios que aplican filtros, exploran segmentos y exportan información.
- **Equipo técnico:** responsables del desarrollo, integración y mantenimiento de la aplicación.

## 2. Requerimientos Funcionales

### 2.1. Fase 1 – Dashboard de consultas y reportes

**Objetivo:** habilitar carga, exploración y visualización de datos con filtros interactivos.

- RF-1. **Carga y validación de datos.** El sistema debe permitir la carga de datasets (CSV u otra fuente), validando estructura, tipos de datos y registros nulos/inconsistentes. Debe generar un reporte de calidad de datos.
- RF-2. **Filtros interactivos.** El usuario podrá aplicar filtros dinámicos sobre múltiples atributos simultáneamente (ej. edad, educación, préstamos, canal de contacto, mes, etc.). Los filtros deben poder combinarse entre sí y deben afectar en tiempo real todos los gráficos y tablas del dashboard.
- RF-3. **Visualización de KPIs y métricas clave.** El sistema debe mostrar indicadores como:
- Tasa de conversión global y segmentada (por campaña, canal, edad) con variación temporal.
  - Número total de contactos realizados.
  - Distribución de clientes por segmentos (edad, educación, préstamos).
  - Duración media de las llamadas, comparando llamadas exitosas vs no exitosas y mostrando su distribución por rangos.

**Tipos de gráficos sugeridos:** Barras y columnas, histogramas, líneas, circulares/donut, heatmaps, boxplots, tablas dinámicas.

- RF-4. **Explorador tabular de datos.** Tabla interactiva con registros filtrados. Debe permitir ordenar, paginar y exportar resultados (CSV/Excel).
- RF-5. **Exportación y reportes consolidados.** Reportes descargables en PDF/Excel con KPIs, gráficos y filtros aplicados.
- RF-6. **Gestión de usuarios y perfiles de acceso.** El sistema debe contemplar la posibilidad de perfiles diferenciados: administrador, analista y lector. Los permisos específicos serán implementados en la **Fase 2**. En la **Fase 1**, el sistema operará con un único nivel de acceso.
- RF-7. **Histórico de consultas y reutilización de filtros.** El sistema debe guardar un historial de consultas y permitir guardar filtros personalizados para reutilizarlos.

### 2.2. Fase 2 – Motor de recomendaciones y soporte a decisiones

**Objetivo:** evolucionar hacia recomendaciones personalizadas y analítica avanzada.

- RF-8. **Segmentación automática de clientes.** Identificación automática de grupos de clientes mediante algoritmos de segmentación.
- RF-9. **Motor de recomendación de productos.** Generación de recomendaciones personalizadas (créditos, hipotecas, depósitos) con justificación basada en atributos.

- RF-10. **KPIs avanzados de decisión.** Cálculo de métricas como probabilidad de aceptación por segmento y rentabilidad proyectada.
- RF-11. **Simulación de escenarios (“what-if analysis”).** El usuario podrá simular cambios de parámetros y visualizar el impacto estimado en conversión y rentabilidad.
- RF-12. **Reportes ejecutivos con insights.** Generación automática de reportes ejecutivos con segmentación, recomendaciones, KPIs clave y escenarios comparativos.

**Tipos de gráficos sugeridos para esta fase:** Barras apiladas, dispersión (scatter plots), curvas ROC/precisión-recall, diagramas de Sankey, clustering plots.

### 3. Requerimientos No Funcionales

- **Usabilidad:** interfaz clara, filtros intuitivos, visualizaciones responsivas.
- **Escalabilidad:** arquitectura por componentes para añadir nuevos productos y fuentes de datos.
- **Seguridad:** control de accesos por roles (definidos en Fase 2) y encriptación de datos sensibles.
- **Mantenibilidad:** código modular y documentación actualizada.
- **Disponibilidad:** acceso web con soporte multiplataforma.

### 4. Plan de Iteraciones

- **Iteración 1 (Fase 1):** carga de datos, filtros múltiples, KPIs básicos, explorador tabular, exportación e historial de filtros.
- **Iteración 2 (Fase 2):** segmentación, motor de recomendaciones, KPIs avanzados, simulación de escenarios, reportes ejecutivos y control de accesos por roles.

### 5. Glosario

- **Tasa de conversión:** porcentaje de clientes que aceptan un producto.
- **Segmentación:** agrupación de clientes con atributos similares.
- **Recomendación:** sugerencia de producto financiero personalizada.

### 6. Fuentes bibliográficas

#### Referencias

- [1] Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62, 22–31. <https://doi.org/10.1016/j.dss.2014.03.001>

- [2] Moro, S., Cortez, P., & Rita, P. (2011). Using data mining for bank direct marketing: An application of the CRISP-DM methodology. In *Proceedings of the European Simulation and Modelling Conference* (pp. 117–121). EUROSIS-ETI.
- [3] UCI Machine Learning Repository. (2012). *Bank Marketing Data Set*. University of California, Irvine, School of Information and Computer Sciences. Disponible en: <https://archive.ics.uci.edu/dataset/222/bank+marketing>
- [4] 1010data. (2020). *Bank Marketing Data Set — Field Descriptions*. 1010data Machine Learning Examples. Disponible en: <https://docs.1010data.com/MachineLearningExamples/BankMarketingDataSet.html>

## Anexo A — Diccionario de Datos

Campo	Tipo	Descripción	Valores válidos / Rango
age	Entero	Edad del cliente.	Min=17, Max=98.
job	Categórico	Ocupación del cliente.	admin., blue-collar, entrepreneur, housemaid, management, retired, self-employed, services, student, technician, unemployed, unknown.
marital	Categórico	Estado civil.	divorced, married, single, unknown.
education	Categórico	Nivel educativo alcanzado.	basic.4y, basic.6y, basic.9y, high.school, illiterate, professional.course, university.degree, unknown.
default	Binario/categ.	¿Crédito en mora?	yes, no, unknown.
housing	Binario/categ.	¿Préstamo hipotecario?	yes, no, unknown.
loan	Binario/categ.	¿Préstamo personal?	yes, no, unknown.
contact	Categórico	Canal/tipo de contacto.	cellular, telephone.
month	Categórico	Mes del último contacto.	jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec.
day_of_week	Categórico	Día de la semana del contacto.	mon, tue, wed, thu, fri.
duration	Entero (s)	Duración del último contacto (segundos). <i>Nota: no está disponible antes de la llamada.</i>	Min=0, Max=4918.
campaign	Entero	Nº de contactos durante esta campaña (incluye el último).	Min=1, Max=56.
pdays	Entero	Días desde el contacto previo (si lo hubo).	Min=0, Max=999.
previous	Entero	Nº de contactos previos antes de esta campaña.	Min=0, Max=7.
poutcome	Categórico	Resultado de la campaña previa.	failure, nonexistent, success.
emp.var.rate	Real	Variación del empleo (trimestral).	Min=-3.4, Max=1.4.
cons.price.idx	Real	Índice de precios al consumidor (mensual).	Min=92.201, Max=94.767.
cons.conf.idx	Real	Índice de confianza del consumidor (mensual).	Min=-50.8, Max=-26.9.
euribor3m	Real	Tasa Euribor a 3 meses (diaria).	Min=0.634, Max=5.045.

*Continúa en la siguiente página*

Campo	Tipo	Descripción	Valores válidos / Rango
nr.employed	Real	Nº de empleados (trimestral).	Min=4963.6, Max=5228.1.
y	Binario	¿Suscribió depósito a plazo?	yes, no.

**Notas:**

- Valores y rangos numéricos calculados sobre `bank-additional-full.csv` (41,188 filas, 21 columnas).
- Algunas variables categóricas incluyen `unknown` para denotar información no disponible.
- La versión clásica `bank-full.csv` (45,211 filas) difiere en columnas; ajustar si se usa esa fuente.

## Anexo B — Mapeo de Filtros a Visualizaciones

Filtro / Variable	Tipo de dato	Visualización sugerida
age	Numérica continua	Histograma de distribución; Boxplot por segmentos (ej. y).
job	Categórica nominal	Barras (conteo o tasa de conversión por ocupación).
marital	Categórica nominal	Circular/donut (proporciones por estado civil).
education	Categórica ordinal	Barras ordenadas (tasa por nivel educativo).
default	Binaria/categórica	Barras comparativas (sí/no).
housing	Binaria/categórica	Barras comparativas (sí/no).
loan	Binaria/categórica	Barras comparativas (sí/no).
contact	Categórica nominal	Barras (celular vs teléfono).
month	Categórica temporal	Líneas (tendencia mensual de contactos o conversión).
day_of_week	Categórica temporal	Heatmap (día de la semana vs tasa de conversión).
duration	Numérica continua	Histograma; boxplot comparativo (y = yes/no).
campaign	Numérica discreta	Histograma; dispersión (campaign vs conversión).
pdays	Numérica discreta	Histograma; dispersión vs tasa de conversión.
previous	Numérica discreta	Barras (nº de contactos previos vs conversión).
poutcome	Categórica nominal	Barras apiladas (resultado previo vs aceptación).
emp.var.rate	Numérica continua	Serie temporal o dispersión (relación con conversión).
cons.price.idx	Numérica continua	Serie temporal o dispersión.
cons.conf.idx	Numérica continua	Serie temporal; dispersión (comparación con aceptación).
euribor3m	Numérica continua	Líneas (variación diaria); dispersión vs aceptación.
nr.employed	Numérica continua	Serie temporal; dispersión (relación con conversión).
y (objetivo)	Binaria	Circular (yes/no); tarjeta KPI de tasa de conversión.

### Notas:

- Variables numéricas: histogramas y boxplots; combinar con y para ver efecto en conversión.
- Categóricas: priorizar barras y proporciones.
- Temporales (month, day\_of\_week): líneas y mapas de calor para estacionalidad.

## Anexo C — Definición Formal de KPIs

KPI	Definición	Fórmula matemática
<b>Tasa de conversión (CR)</b>	Porcentaje de clientes contactados que suscriben un depósito a plazo fijo.	$CR = \frac{N_{y=yes}}{N_{total}} \times 100$
<b>Número total de contactos</b>	Total de llamadas realizadas en la campaña.	$N_{contactos} = \sum_{i=1}^n 1$
<b>Duración promedio de llamadas</b>	Tiempo medio de duración de los contactos realizados.	$\bar{D} = \frac{\sum_{i=1}^n duration_i}{N_{total}}$
<b>Tasa de éxito por canal</b>	Proporción de conversiones por canal (cellular, telephone).	$CR_{canal=j} = \frac{N_{y=yes, contact=j}}{N_{contact=j}} \times 100$
<b>Conversión por segmento de edad</b>	Tasa de conversión por rangos de edad.	$CR_{[a,b]} = \frac{N_{y=yes, age \in [a,b]}}{N_{age \in [a,b]}} \times 100$
<b>Impacto del historial previo</b>	Conversión condicionada a poutcome.	$CR_{poutcome=k} = \frac{N_{y=yes, poutcome=k}}{N_{poutcome=k}} \times 100$
<b>Índice de eficiencia por campaña</b>	Conversión relativa a intentos de contacto.	$Eff = \frac{N_{y=yes}}{\sum_{i=1}^n campaign_i}$
<b>Rentabilidad proyectada (simplificada)</b>	Retorno estimado con ganancia unitaria $G$ y costo por llamada $C$ .	$R = N_{y=yes} \cdot G - N_{total} \cdot C$

### Notas:

- $N_{y=yes}$ : número de clientes que aceptaron el producto.
- $N_{total}$ : número total de contactos realizados.
- Las fórmulas pueden condicionarse por filtros (ocupación, educación, préstamos, etc.).



## Anexo C-2 — Mapa de KPIs por Fase y Vista Recomendada

KPI	Descripción corta (referencia a fórmula en Anexo C)	Fase	Vista recomendada
Tasa de conversión (CR)	$CR = \frac{N_{y=yes}}{N_{total}} \times 100.$	1	Tarjeta KPI; barras por segmento; línea temporal.
Número total de contactos	$N_{contactos} = \sum 1.$	1	Tarjeta KPI; barras (por mes, ocupación).
Duración promedio de llamadas	$\overline{D} = \frac{\sum duration_i}{N_{total}}.$	1	Tarjeta KPI; histograma; boxplot (y = yes/no).
Tasa de éxito por canal	$CR_{canal=j} = \frac{N_{y=yes, contact=j}}{N_{contact=j}} \times 100.$	1	Barras comparativas (cellular vs telephone).
Conversión por segmento de edad	$CR_{[a,b]} = \frac{N_{y=yes, age \in [a,b]}}{N_{age \in [a,b]}} \times 100.$	1	Barras horizontales por rangos; heatmap con educación.
Impacto del historial previo	$CR_{poutcome=k} = \frac{N_{y=yes, poutcome=k}}{N_{poutcome=k}} \times 100.$	1	Barras apiladas (poutcome vs aceptación).
Índice de eficiencia por campaña	$Eff = \frac{N_{y=yes}}{\sum campaign_i}.$	1	Dispersión (campaign vs CR); líneas comparativas.
Rentabilidad proyectada (simplificada)	$R = N_{y=yes} \cdot G - N_{total} \cdot C.$	2	Tarjeta KPI; columnas (escenarios “what-if”).

### Criterio general:

- Fase 1: lectura operativa (tarjetas KPI, barras, líneas, histogramas).
- Fase 2: métricas estratégicas con escenarios (columnas agrupadas, comparativas).