

Raccolta e analisi dei requisiti per un progetto di  
analisi dati

Analisi, Valutazione del modello di credit scoring  
di Revolut Bank

# Revolut



Giuseppe Schillaci

Epicode Project 1

30 Settembre 2022

<https://github.com/giuschil/epicode-data-analyst/tree/main/project1-week1>

## Descrizione del caso di studio

### Storia

Revolut è una piattaforma finanziaria globale, supportata da un'app che consente in ogni momento di gestire le tue finanza con un semplice smartphone. Revolut Ltd è una società fintech, con sede nel Regno Unito fondata nel 2015. Tuttavia, già nel 2018 ha raggiunto lo status di unicorno. Con ciò si intende che è stata valutata più di un miliardo di dollari. La società ha realizzato un conto corrente multivaluta.

A partire dal 2021 ha ottenuto la licenza bancaria anche in Italia dove viene riconosciuta a tutti gli effetti come una banca, dando quindi accesso ad un vero e proprio conto corrente. Tuttavia, esso non dà IBAN italiano, ma l'estensione LT è riconosciuta nell'area SEPA, cioè dei pagamenti europei. Revolut offre una carta di debito prepagata che consente di prelevare denaro (in valuta locale) in oltre 120 paesi del mondo.

### Caso

Dal 2021 ha iniziato a concedere prestiti e credito al consumo direttamente online ai propri clienti tramite l'app proprietaria.

La suddetta analisi dei requisiti è necessaria per la progettazione di uno strumento per le analisi, valutazione e migioria del modello già utilizzato da Revolut, in particolare si propone di:

1. Migliorare il modello di credit scoring do Revolut Bank per migliorare l'efficacia con cui concede credito ai suoi clienti
2. Poter analizzare e predire tramite dei modelli di ML quali dei suoi clienti potrebbe voler richiedere del credito.
3. Poter analizzare e predire il default rate dei clienti che ha già ottenuto il credito.
4. Predire quali e quando i clienti potrebbero aver necessità di credito per i loro acquisti e fornire i dati al team marketing per migliorare le leads performance

## Analisi e raccolta dei requisiti

Per la creazione del progetto di analisi per la banca Revolut il team Data & Technology della società di consulenza Epicode Consulting ha raccolto i dati seguenti:

Dataset	dimensione dati	formato	link fonte	descrizione
---------	-----------------	---------	------------	-------------

dataset_bank.csv	10 clienti	csv	<a href="https://api.revolut.com/bankdata/">https://api.revolut.com/bankdata/</a>	Dataset della banca sui clienti che hanno richiesto un credito e risultato del credit score
dataset_fiscalagency.json	10 contribuenti	json	<a href="https://api.agenziaentrate.gov.it/portale/">https://api.agenziaentrate.gov.it/portale/</a>	Dataset dell'agenzia delle entrate sui dieci clienti che hanno richiesto il credito
dataset_transactions.xml	10 transazioni di 1 cliente	xml	<a href="https://api.stripe.com/bank/37475852">https://api.stripe.com/bank/37475852</a>	Dataset creato da azienda di credito terza su tutti i pagamenti online e offline effettuati

## Requisiti per la Banca

Solo i già clienti possono richiedere un prestito o credito al consumo.

I clienti possono richiedere un credito della durata massima 30 anni, di quantità non superiore a 300.000 €, non è importante sapere il tasso a cui è stato richiesto il prestito.

I clienti possono chiedere un credito per l'acquisto di una car, house, other

## Requisiti per le Transazioni

I dati sulle transazioni sono i dati raccolti dall'istituto di credito online per i pagamenti (esempio Stripe) trasmessi tramite connessione API.

I dati sono in formato XML e racco

## Requisiti per i dati dell'Agenzia delle entrate

I dati dell'agenzia dell'entrate sono rilasciati in formato json sono dati protetti e contengono l'anagrafica del contribuente

## Team necessario

- Data Engineer
  - Progetta e struttura il Data Lake di dati non strutturati su server distribuiti tramite Hadoop e linguaggi di programmazione Java e Sql.
- Data Analyst
  - Analizza i dati tramite interrogazioni della base di dati, crea report e dashboard sulle spese maggiormente fatte, la media degli importi spesi, i luoghi e la categoria merceologica maggiormente acquistata.
- Data scientist
  - Crea il modello Crea un modello per predire quale cliente potrebbe, in base alla sua capacità di spesa, voler richiedere un prestito, anticipando così le campagne di email marketing

## Strutturazione dei requisiti in gruppi di frasi omogenee

### Utenti

Per gli utenti (10.000), identificati da un codice numerico univoco, il nome, il cognome, l'età, il genere, la città, la nazione e lo stato civile.

### Strutture

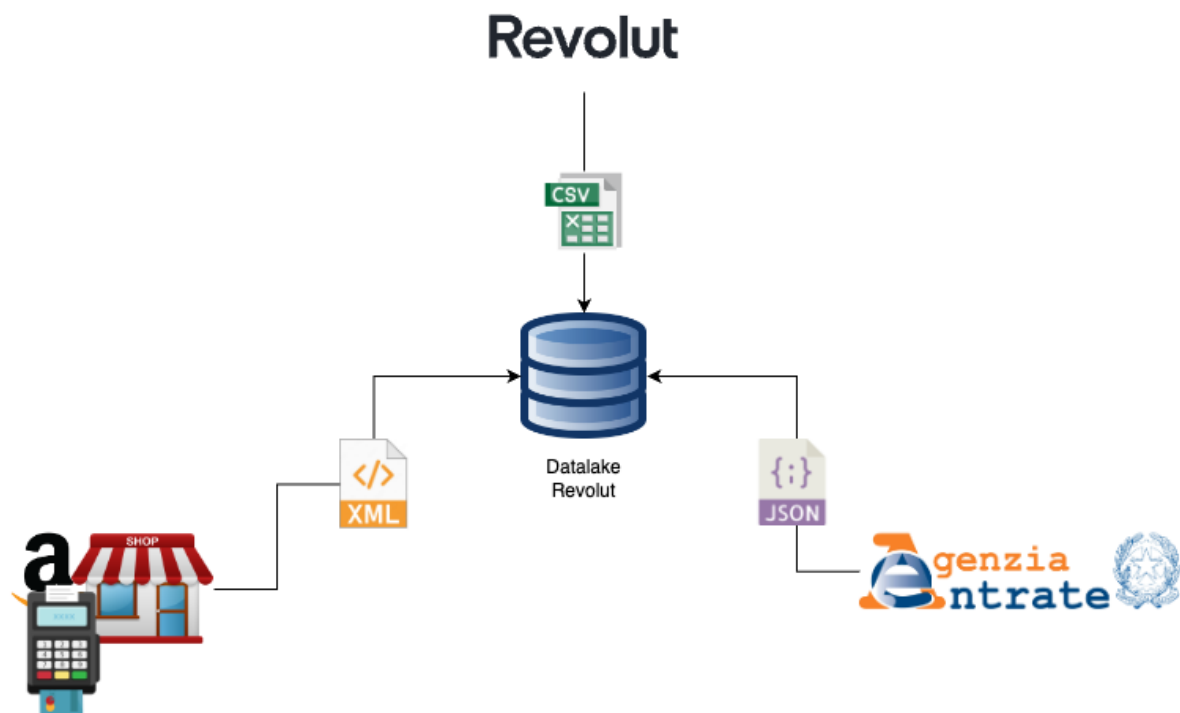
Per le strutture (1000), identificate da un codice numerico univoco, il nome della struttura, il tipo della struttura, la città, il prezzo e la foto della struttura.

### Prenotazioni

Le prenotazioni sono identificate da un codice numerico univoco, una data di inizio prenotazione, una data di fine prenotazione, l'importo della prenotazione e il numero degli ospiti.

## Glossario dei termini

Termine	Descrizione	Sinonimi	Collegamenti
Cliente	Cliente della banca che richiede il credito	Utente	
Prestito	Somma di denaro concessa al cliente dalla banca	Credito, Credito al consumo	
Banca	Azienda che permette la concessione di denaro ai suoi clienti o la possibilità di effettuare pagamenti	Istituto Bancario, Azienda di Credito	Utenti, Struttura
Agenzia delle entrate	L'Agenzia delle entrate è un'agenzia fiscale della pubblica amministrazione italiana svolge le funzioni relative ad accertamenti e controlli fiscali e alla gestione dei tributi.	Agenzia fiscale	
Transazione	Transazione di denaro effettuata dal cliente della suddetta banca	Pagamento	



## Tipologia di dati

### Dati Revolut Bank

Tipo di dataset strutturato in formato csv fornito dalla banca Revolut

id	numerico	valori	esempio
fiscal_code	varchar		
bank_account_number	varchar		843721

age	number		23
sex	boolean	M,F	
personal_status	varchar	celibe,sposato,nubile	
job	varchar	insegnante,operaio, avvocato	
housing	boolean	proprietà, affitto	
saving_account	number		
credit_amount	number		
duration	number		
purpose	varchar	car,house,other	
decision	boolean	Yes,No	

## Dati transazioni-pagamenti online e offline

Tipo di dataset non strutturato in formato xml fornito dal sistema di pagamenti Online

transaction id		
id_struttura_prenotata		
category		
date		
time		
company		
url		
title		

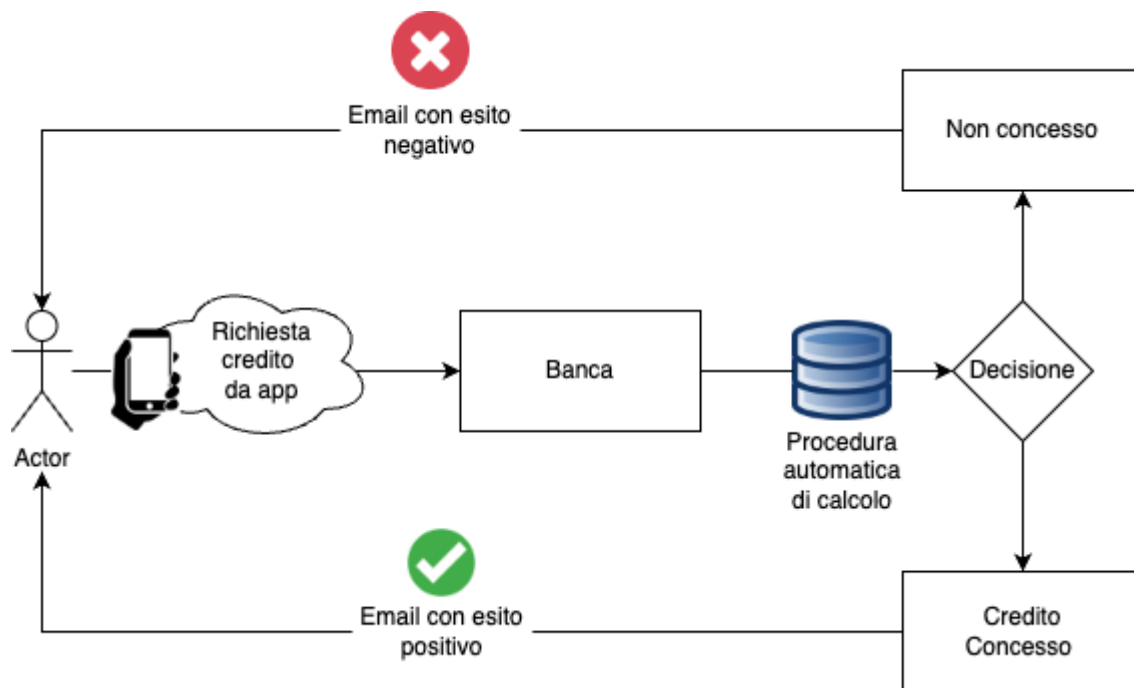
price		
currency		
location		
city		
street		

## Dati Agenzia delle entrate

Tipo di dataset non strutturato in formato json fornito dall'agenzia delle entrate

codice		
nome		
cognome		
cod_fiscale		
indirizzo		
citta		
cap		
numero		
tipo		
nome_zona		
reddito		
reddito dipendente		
reddito altri		
persone_a_carico		

## Supporto alle decisioni aziendali



### Analisi di business

- Trend
  - Analisi dei periodi storici
- Peak seasons
  - Periodo preferito dagli utenti
- Peak location
  - Quali Città hanno avuto più prenotazioni
  - Quali tipi di strutture sono state maggiormente prenotate
- Target customers
  - Quali tipologie di utente ha effettuato più prenotazioni
- Campagne Marketing/Advertising
  - Consigliare le strutture di quale città promuovere con campagne di advertising. Sapere l'effetto delle campagne marketing effettuate prima e dopo la data delle campagne

### Analisi predittiva

- Trend: Previsione nei prossimi 3 mesi
- Peak seasons
  - Periodo preferito dagli utenti
- Peak location
  - Quali Città saranno le preferite dagli utenti nei prossimi 3 mesi
- Target customers
  - Quali tipologie di utente effettuerà più prenotazioni nei prossimi 3 mesi



- Campagne Marketing/Advertising
  - Consigliare le strutture di quale città promuovere con campagne di advertising
  - Sapere l'effetto delle campagne marketing effettuate prima e dopo la data delle campagne

## Storia/Infografica

1. Infografica sul cliente tipo per le città più grandi.
2. Infografica su tipo di prenotazione maggiormente effettuata
  - a. Venezia uomo sposato, sopra i 40, per san valentino;
  - b. Quanti uomini sposati hanno prenotato a san valentino a venezia;
  - c. Milano, uomini e donne under 35, nel periodo della fashion week.
3. Infografica su città con più prenotazioni (immagine sotto)

## Outcome

## Caso d'uso

Interrogazione della base di dati delle tabelle utenti e prenotazioni. In particolare ID\_prenotazione, ID\_utente, raggruppati per età (18-25, 25-34, 35-44, 45-60). Il risultato sarà un grafico di tipo istogramma da cui possiamo identificare che il maggior numero di utenti appartiene alla fascia [45-60], seguita dalla fascia [25-34].

I dati sono da moltiplicare per 1000 i dati sono riferiti all'ultimo anno 2021.

## Analisi Proposte

Sono richieste le seguenti analisi:

- Analisi statistica della concessione del credito ai clienti raggruppati per classe di età (18-25, 25-34, 35-44, 45-60).
- Analisi sul numero dei prestiti concessi

## Possibili implementazioni future

- 1) Anticipare la richiesta di credito da parte degli utenti tramite la creazione di modelli predittivi.
- 2) Migliorare il modello per evitare il default rate dei clienti
- 3)