



# Lavori di Gruppo

Anno Accademico 2025/2026

Analisi Dati Applicate a Diversi Settori

**Corso:** Coding e Big Data

**Docente:** Vincenzo Nardelli

**Email:** [vincenzo.nardelli@unicatt.it](mailto:vincenzo.nardelli@unicatt.it)

## Regole e Modalità di Consegna

### Obiettivi del Progetto

Ogni team dovrà:

1. **Analisi dei dati:** Utilizzare i dati forniti e tutti i codici e modelli visti durante il corso per raggiungere l'obiettivo specifico del proprio team
2. **Documentazione:** Produrre un documento che contenga:
  - Overview del contesto in cui si colloca il progetto
  - Descrizione dettagliata di tutte le scelte metodologiche effettuate
  - Analisi svolte e risultati ottenuti
  - Interpretazione dei risultati e conclusioni
3. **Riflessione critica:** Includere considerazioni sull'impatto dei Big Data e idee future nel settore assegnato (sanità, finanza, marketing, logistica, retail, assicurazioni, immobiliare). In particolare, rispondere alle seguenti domande:
  - Quali opportunità e rischi ci saranno grazie ai Big Data nel vostro settore?
  - Oltre all'algoritmo e l'analisi dei dati proposti dal gruppo, cosa si potrebbe fare per risolvere il problema?
  - Quali sviluppi futuri sono possibili in questo ambito?
4. **Presentazione:** Presentare il progetto durante la lezione del **12 dicembre 2025**

### Scadenze

- **10 dicembre 2025:** Consegna codice e documentazione
- **12 dicembre 2025:** Consegna presentazione e presentazione in aula

### Strumenti e Risorse Consentite

#### È consentito:

- Utilizzare codici e metodologie visti durante le lezioni del corso
- Utilizzare sistemi di intelligenza artificiale (es. ChatGPT, GitHub Copilot, etc.)
- Utilizzare codici, pacchetti o strumenti di analisi diversi da quelli visti a lezione

#### Requisiti obbligatori:

**Importante:** ⚠️ Qualsiasi strumento, pacchetto o metodologia non vista a lezione deve essere:

- **Dichiarato esplicitamente** nella documentazione
- **Compreso nel suo funzionamento** dal team

## Modalità di Valutazione

Durante la presentazione del 12 dicembre 2025:

- Verranno poste domande sul progetto e sui risultati ottenuti
- Verranno poste domande sugli strumenti e metodologie utilizzate
- Sarà valutata la comprensione degli approcci adottati
- Sarà valutata la capacità di interpretare e comunicare i risultati

*Buon lavoro a tutti i team!*

# 1 Finanza: Rischio Creditizio

## Contesto

Nel settore fintech, l'analisi dei dati è fondamentale per identificare i clienti a rischio di insolvenza e per sviluppare strategie efficaci di gestione del rischio. Attraverso l'analisi dei dati, le aziende possono prevedere comportamenti finanziari, personalizzare le offerte e migliorare la gestione del credito. Ad esempio, molte istituzioni finanziarie utilizzano modelli predittivi basati su dati demografici e comportamentali per valutare la probabilità di default dei clienti.

In qualità di data scientist nel reparto Risk Management di una startup fintech, il vostro compito è esaminare i dati dei clienti per identificare i fattori che contribuiscono alle difficoltà di pagamento. L'obiettivo è sviluppare modelli predittivi che consentano di anticipare il rischio di default e implementare misure preventive. Questo approccio proattivo è essenziale per mantenere la stabilità finanziaria dell'azienda e offrire soluzioni adeguate ai clienti. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che influenzano la solvibilità dei clienti, permettendo all'azienda di adottare strategie mirate per minimizzare il rischio di default.

## Obiettivo

Identificare i fattori chiave che influenzano il rischio di default dei clienti, al fine di implementare strategie efficaci per minimizzare tale rischio.

## Informazioni disponibili:

Variabile	Descrizione
default	Cliente con difficoltà di pagamento (TARGET) – Si/No
genere	Genere del cliente – Maschio/Femmina
possiede_auto	Possiede automobile – Si/No
possiede_immobile	Possiede immobile – Si/No
numero_figli	Numero di figli
reddito_totale	Reddito totale annuale
importo_credito	Importo del credito richiesto
tipo_reddito	Tipo di reddito – Lavoratore, Commerciale, Pensionato, Disoccupato
livello_istruzione	Livello di istruzione – Istruzione Superiore, Secondaria, Laurea
stato_famiglia	Stato familiare – Sposato, Single, Unione Civile, Vedovo, Separato
tipo_abitazione	Tipo di abitazione – Casa/Appartamento, Con Genitori, Appartamento Comunale
tipo_occupazione	Tipo di occupazione – Operai, Personale Base, Manager, Tecnici Specializzati

## 2 Logistica: Ottimizzazione Consegne

### Contesto

Nel settore dell'e-commerce, la puntualità delle consegne è cruciale per garantire la soddisfazione del cliente e mantenere un vantaggio competitivo. L'analisi dei dati consente di identificare i fattori che influenzano i ritardi nelle spedizioni, permettendo all'azienda di implementare strategie mirate per migliorare l'efficienza logistica. Ad esempio, molte aziende utilizzano modelli predittivi per anticipare possibili ritardi e ottimizzare le risorse in base alle esigenze operative.

In qualità di data analyst nel reparto Operations di un'azienda internazionale di e-commerce specializzata nella vendita di prodotti elettronici, il vostro compito è esaminare i dati relativi alle spedizioni per identificare le variabili che incidono maggiormente sui ritardi nelle consegne. Attraverso questa analisi, potrete sviluppare modelli predittivi che consentano di anticipare i ritardi e implementare misure preventive. Questo approccio proattivo è essenziale per mantenere la soddisfazione del cliente e ottimizzare le operazioni logistiche. L'analisi dei dati vi permetterà di individuare pattern e tendenze che influenzano la puntualità delle consegne, fornendo all'azienda le informazioni necessarie per adottare strategie efficaci.

### Obiettivo

Identificare i fattori chiave che influenzano il ritardo delle consegne, al fine di implementare strategie efficaci per migliorare l'efficienza logistica e la soddisfazione del cliente.

### Informazioni disponibili:

Variabile	Descrizione
id_spedizione	Identificativo univoco della spedizione
modalita_spedizione	Modalità di spedizione – Nave, Aereo, Strada
valutazione_cliente	Valutazione del cliente (1 = peggiore, 5 = migliore)
costo_prodotto	Costo del prodotto in dollari statunitensi
acquisti_precedenti	Numero di acquisti precedenti del cliente
importanza_prodotto	Importanza del prodotto – bassa, media, alta
genere	Genere del cliente – Maschio, Femmina
sconto_offerto	Percentuale di sconto offerto sul prodotto (0-100)
peso_grammi	Peso del prodotto in grammi
consegna_puntuale	Consegna puntuale (TARGET) – Si/No (Si = puntuale, No = in ritardo)

### 3 Marketing: Campagne Promozionali

#### Contesto

Nel settore della grande distribuzione, comprendere il comportamento dei clienti è fondamentale per sviluppare strategie di marketing efficaci. L'analisi dei dati consente di identificare i segmenti di clientela più propensi a rispondere positivamente alle offerte promozionali, ottimizzando così le risorse e massimizzando il ritorno sugli investimenti. Ad esempio, molte aziende utilizzano modelli predittivi basati su dati demografici e comportamentali per personalizzare le campagne e aumentare il tasso di risposta.

In qualità di data scientist nel reparto Marketing Analytics di un'azienda leader nella grande distribuzione, il vostro compito è esaminare i dati dei clienti per identificare i fattori che influenzano la probabilità di risposta alle offerte promozionali. Attraverso questa analisi, potrete sviluppare modelli predittivi che consentano di anticipare le reazioni dei clienti e implementare strategie di marketing mirate. Questo approccio proattivo è essenziale per migliorare l'efficacia delle campagne promozionali e aumentare la soddisfazione del cliente. L'analisi dei dati vi permetterà di individuare pattern e tendenze che influenzano le decisioni d'acquisto, fornendo all'azienda le informazioni necessarie per adottare strategie efficaci.

#### Obiettivo

Identificare i fattori chiave che influenzano la probabilità di risposta positiva alle offerte promozionali, al fine di implementare strategie efficaci per migliorare l'efficacia delle campagne di marketing.

#### Informazioni disponibili:

Variabile	Descrizione
anno_nascita	Anno di nascita del cliente
stato_civile	Stato civile del cliente – Sposato, Single, Convivente, Divorziato, Vedovo
reddito	Reddito annuale del cliente
figli_piccoli_casa	Numero di bambini piccoli nel nucleo familiare
adolescenti_casa	Numero di adolescenti nel nucleo familiare
spesa_vini	Spesa per vini negli ultimi due anni
spesa_frutta	Spesa per frutta negli ultimi due anni
spesa_carne	Spesa per prodotti a base di carne negli ultimi due anni
spesa_pesce	Spesa per prodotti ittici negli ultimi due anni
spesa_dolci	Spesa per dolci negli ultimi due anni
spesa_oro	Spesa per prodotti premium negli ultimi due anni
risposta_campagna	Risposta alla campagna (TARGET) – Sì/No

## 4 Sanità: Prevenzione Malattie Cardiache

### Contesto

Nel settore sanitario, l'analisi dei dati è fondamentale per identificare i fattori di rischio associati alle malattie cardiache e sviluppare strategie efficaci di prevenzione e intervento. Attraverso l'analisi dei dati clinici e degli stili di vita, le organizzazioni sanitarie possono prevedere quali pazienti sono a maggior rischio, personalizzare i programmi di prevenzione e migliorare gli esiti clinici. Ad esempio, molte strutture sanitarie utilizzano modelli predittivi basati su indicatori di salute, comportamenti e caratteristiche demografiche per identificare i pazienti ad alto rischio e implementare interventi mirati.

In qualità di data analyst nel reparto Health Analytics di un'azienda sanitaria, il vostro compito è esaminare i dati dei pazienti per identificare i fattori che contribuiscono al rischio di malattie cardiache. L'obiettivo è sviluppare modelli predittivi che consentano di anticipare quali pazienti sono a maggior rischio di sviluppare malattie coronariche o infarti del miocardio e implementare programmi di prevenzione personalizzati. Questo approccio proattivo è essenziale per migliorare la salute della popolazione e ridurre i costi sanitari. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che influenzano la salute cardiovascolare, permettendo all'azienda di adottare strategie preventive più efficaci.

### Obiettivo

Identificare i fattori chiave che influenzano il rischio di malattie cardiache, al fine di implementare strategie efficaci per la prevenzione e l'intervento precoce.

### Informazioni disponibili:

Variabile	Descrizione
malattia_cardiaca	Presenza malattia cardiaca (TARGET) – Si/No
indice_massa_corporea	BMI (Body Mass Index) – range 12-95
fumatore	È fumatore – Si/No
consumo_alcol	Consumo eccessivo di alcol – Si/No
ictus	Ha avuto ictus – Si/No
salute_fisica	Giorni di cattiva salute fisica negli ultimi 30 giorni (0-30)
salute_mentale	Giorni di cattiva salute mentale negli ultimi 30 giorni (0-30)
difficolta_camminare	Difficoltà nel camminare – Si/No
sex	Sesso – Maschio/Femmina
categoria_eta	Fascia di età – 18-24, 25-29, ..., 80 o più
etnia	Etnia – Bianco, Nero, Ispanico, Asiatico, Altro, Nativo Americano
diabetico	Stato diabete – Si, No, No (Prediabete), Si (Gravidanza)
attivita_fisica	Pratica attività fisica – Si/No
salute_generale	Percezione salute generale – Eccellente, Molto Buona, Buona, Discreta, Scarsa
ore_sonno	Ore di sonno medie per notte (1-24)
asma	Ha asma – Si/No
malattia_renale	Ha malattia renale – Si/No
cancro_pelle	Ha cancro della pelle – Si/No

## 5 Retail Moda: Performance Negozi

### Contesto

Nel settore della moda retail, l'analisi dei dati è fondamentale per ottimizzare le performance dei negozi e massimizzare la redditività. Attraverso l'analisi dei dati operativi, le aziende possono identificare i fattori che contribuiscono al raggiungimento degli obiettivi di vendita, ottimizzare la gestione del personale e dell'inventario, e migliorare l'efficienza complessiva. Ad esempio, molte catene di negozi utilizzano modelli predittivi basati su dati di organico, inventario e caratteristiche del punto vendita per prevedere le performance e allocare le risorse in modo ottimale.

In qualità di data scientist nel reparto Retail Analytics di una catena di negozi di moda, il vostro compito è esaminare i dati dei punti vendita per identificare i fattori che contribuiscono al raggiungimento degli obiettivi di vendita. L'obiettivo è sviluppare modelli predittivi che consentano di anticipare quali negozi raggiungeranno i target e implementare strategie di supporto per quelli in difficoltà. Questo approccio proattivo è essenziale per mantenere la redditività dell'azienda e ottimizzare l'allocazione delle risorse. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che influenzano le performance dei negozi, permettendo all'azienda di adottare strategie mirate per massimizzare le vendite.

### Obiettivo

Identificare i fattori chiave che influenzano il raggiungimento degli obiettivi di vendita dei negozi, al fine di implementare strategie efficaci per ottimizzare le performance e l'allocazione delle risorse.

### Informazioni disponibili:

Variabile	Descrizione
obiettivo_vendite	Raggiungimento obiettivo vendite (TARGET) – sì/no
margin	Margine di profitto lordo (gross-profit-margin)
numero_proprietari	Numero di proprietari/manager del negozio
numero_tempo_pieno	Numero di dipendenti a tempo pieno
numero_part_time	Numero di dipendenti part-time
numero_ausiliari	Numero di lavoratori temporanei/aiutanti
ore_settimanali	Ore totali lavorate
ore_per_settimana	Ore lavorate per lavoratore
investimento_locali	Investimento nei locali del negozio
investimento_automazione	Investimento in automazione
dimensione_negozi	Superficie area vendita del negozio (m <sup>2</sup> )



## 6 Assicurazioni: Cross-Selling Prodotti

### Contesto

Nel settore assicurativo, l'analisi dei dati è cruciale per identificare potenziali clienti interessati a nuovi prodotti assicurativi e ottimizzare le strategie di cross-selling. Attraverso l'analisi dei dati demografici e del portafoglio assicurativo esistente, le compagnie possono prevedere quali clienti sono più propensi ad acquistare nuove polizze, personalizzare le offerte e migliorare il tasso di conversione. Ad esempio, molte compagnie assicurative utilizzano modelli predittivi basati su età, reddito, istruzione e polizze già possedute per identificare i clienti target per campagne di marketing mirate.

In qualità di data scientist nel reparto Customer Analytics di una compagnia assicurativa olandese, il vostro compito è esaminare i dati dei clienti per identificare i fattori che influenzano la probabilità di acquisto di una polizza caravan (roulotte). L'obiettivo è sviluppare modelli predittivi che consentano di anticipare quali clienti sono più propensi all'acquisto e implementare campagne di marketing mirate. Questo approccio proattivo è essenziale per ottimizzare il budget marketing e massimizzare il ritorno sugli investimenti. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che caratterizzano i clienti interessati a questo tipo di polizza, permettendo all'azienda di adottare strategie di targeting più efficaci.

### Obiettivo

Identificare i fattori chiave che influenzano la probabilità di acquisto di una polizza caravan, al fine di implementare strategie efficaci per ottimizzare le campagne di cross-selling e massimizzare il tasso di conversione.

### Informazioni disponibili:

Variabile	Descrizione
acquisto	Acquisto polizza caravan (TARGET) – Si/No
eta_media	Età media del capofamiglia (scala 1-6)
reddito_medio	Reddito medio dell'area (scala 1-9)
classe_acquisto	Classe di potere d'acquisto (scala 1-8)
istruzione_alta	Percentuale con istruzione superiore (scala 0-9)
tipo_cliente	Tipo di cliente/segmento (scala 1-41)
contributo_polizza_auto	Contributo polizza auto (scala 0-9)
contributo_polizza_incendio	Contributo polizza incendio (scala 0-9)
contributo_polizza_vita	Contributo polizza vita (scala 0-9)
fascia_eta	Fascia di età – Giovane, Adulto, Senior
fascia_reddito	Fascia di reddito – Basso, Medio, Alto
potere_acquisto	Livello potere d'acquisto – Limitato, Moderato, Elevato
ha_assicurazioni	Possiede già assicurazioni – Si/No

## 7 Immobiliare: Valutazione Proprietà

### Contesto

Nel settore immobiliare, l'analisi dei dati è fondamentale per comprendere le dinamiche del mercato e identificare le caratteristiche che rendono un immobile più desiderabile e di valore. Attraverso l'analisi delle caratteristiche degli immobili e della loro localizzazione, le agenzie possono prevedere i prezzi di mercato, identificare le aree preferite dai clienti e ottimizzare le strategie di pricing e marketing. Ad esempio, molte agenzie immobiliari utilizzano modelli predittivi basati su caratteristiche strutturali, comfort e localizzazione per valutare gli immobili e identificare le opportunità di investimento.

In qualità di data analyst nel reparto Market Intelligence di una grande agenzia immobiliare, il vostro compito è esaminare i dati degli immobili per identificare i fattori che determinano se una proprietà si trova in un'area preferita (area desiderabile con maggiore domanda). L'obiettivo è sviluppare modelli predittivi che consentano di anticipare quali caratteristiche sono associate alle aree preferite e supportare le decisioni di pricing e investimento. Questo approccio proattivo è essenziale per ottimizzare il portafoglio immobiliare e massimizzare il valore per i clienti. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che caratterizzano gli immobili nelle aree più desiderabili, permettendo all'azienda di adottare strategie di valutazione e marketing più efficaci.

### Obiettivo

Identificare i fattori chiave che caratterizzano gli immobili situati in aree preferite, al fine di implementare strategie efficaci per la valutazione immobiliare e l'ottimizzazione del portafoglio.

### Informazioni disponibili:

Variabile	Descrizione
area_preferita	In area preferita (TARGET) – si/no
prezzo	Prezzo di vendita in dollari
dimensione_lotto	Dimensione del lotto in piedi quadrati
camere_letto	Numero di camere da letto
bagni	Numero di bagni
piani	Numero di piani
posti_garage	Posti auto in garage
violetto	Presenza vialetto – si/no
sala_ricreativa	Presenza sala ricreativa – si/no
seminterrato_completo	Seminterrato completo – si/no
riscaldamento_gas	Riscaldamento a gas – si/no
aria_condizionata	Aria condizionata – si/no

## 8 Retail: Acquisizione Clienti

### Contesto

Nel settore retail, l'analisi dei dati è fondamentale per ottimizzare l'acquisizione di nuovi clienti e massimizzare l'efficacia delle campagne marketing. Attraverso l'analisi dei dati demografici, comportamentali e dei canali di comunicazione, le aziende possono identificare i segmenti di clientela più redditizi, personalizzare le strategie promozionali e ottimizzare l'allocatione del budget marketing. Ad esempio, molte aziende retail utilizzano modelli predittivi basati su caratteristiche sociodemografiche e canali media per identificare i clienti con maggiore potenziale di spesa e ottimizzare il mix di comunicazione.

In qualità di data analyst nel reparto Customer Insights di un'azienda leader nelle vendite retail, il vostro compito è esaminare i dati dei clienti per identificare i fattori che influenzano la spesa totale e l'efficacia delle diverse campagne marketing. L'obiettivo è sviluppare modelli predittivi che consentano di anticipare quali clienti diventeranno clienti ad alta spesa e quali canali di comunicazione sono più efficaci per ciascun segmento. Questo approccio proattivo è essenziale per ottimizzare il ritorno sugli investimenti marketing e massimizzare il valore del cliente nel tempo. Attraverso l'analisi dei dati, potrete individuare pattern e tendenze che influenzano il comportamento d'acquisto, permettendo all'azienda di adottare strategie di acquisizione più mirate ed efficaci.

### Obiettivo

Identificare i fattori chiave che influenzano la spesa dei clienti e l'efficacia delle campagne marketing, al fine di implementare strategie efficaci per ottimizzare l'acquisizione di nuovi clienti ad alto valore.

### Informazioni disponibili:

Variabile	Descrizione
genere	Genere del cliente – Maschio/Femmina
totale_figli	Numero totale di figli del cliente
istruzione	Livello di istruzione – Diploma Superiore, Laurea Triennale, Laurea Magistrale
carta_membro	Tipo di carta fedeltà – Bronzo, Normale, Argento, Oro
occupazione	Tipo di occupazione – Professionista, Manuale Specializzato, Dirigente
proprietario_casa	Possiede casa di proprietà – Sì/No
numero_auto_casa	Numero medio di auto possedute
reddito_annuale_medio	Fascia di reddito annuale – €10K-€30K, €30K-€50K, etc.
numero_figli_casa	Numero di figli che vivono in casa
nome_promozione	Nome della promozione ricevuta
tipo_media	Canale di comunicazione – Coupon in Negozio, Televisione, Posta Massiva
alta_spesa	Cliente con spesa alta (TARGET) – Sì/No