BOOM
AiGeneration

AGENTI GPT

COSA SONO E COME FUNZIONANO

GUIDA COMPLETA: DEEP DIVE 4

LA FUNZIONE RESEARCH & ANALYSIS

SWIPE FOR MORE

BOOM
AiGeneration

Come funziona la funzione Research & Analysis? Vediamo un esempio pratico:

Analisi del rischio bancario

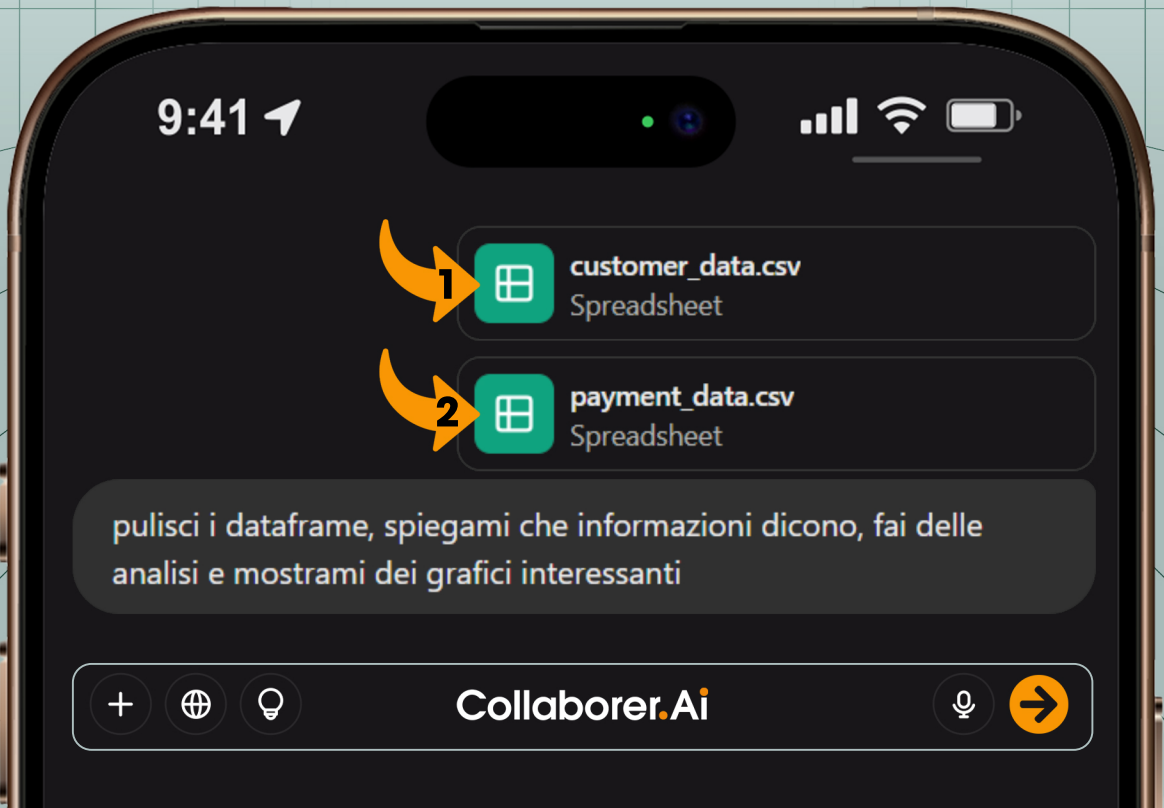
Analizziamo i dati di alcuni clienti per identificare quelli a rischio di insolvenza.

Collaborer.Ai

Per farlo, carichiamo due file contenenti:

1. **Dati demografici dei clienti**
   (età, reddito, storico lavorativo)

2. **Storico delle transazioni e dei pagamenti**
   (frequenza, importi, ritardi nei pagamenti)



L'agente GPT elabora questi file e genera un output strutturato che segue i seguenti passi:

1. Pulizia e verifica dei dati

• Identificazione e gestione di **dati mancanti o anomali** (tramite Pandas, NumPy)

• Standardizzazione dei formati
(es. date con datetime, normalizzazione delle valute)

• Eliminazione dei **dati duplicati** (uso di chiavi univoche, SQL DISTINCT, hashing)

• Conversione di dati non strutturati in un formato analizzabile (es. estrazione da PDF con PyMuPDF o OCR con Tesseract)

Problemi riscontrati nei dati

1. **Customer Data** (`customer_data.csv`)
   - `fea_2` ha **149 valori mancanti**.
   - Alcune feature (es. `fea_7`) hanno valori negativi, da verificare.
2. **Payment Data** (`payment_data.csv`)
   - `prod_limit` ha **6118 valori mancanti**.
   - `highest_balance` ha **409 valori mancanti**.
   - `report_date` ha **1114 valori mancanti**.
   - `update_date` ha **26 valori mancanti**.
   - `new_balance` ha valori molto alti e anche negativi.

Pulizia e gestione dei valori mancanti

- Riempirò `fea_2`, `prod_limit` e `highest_balance` con la mediana, per evitare distorsioni.
- Eliminerò `report_date`, poiché oltre 1000 valori mancanti potrebbero renderla poco utile.
- Verificherò i valori negativi di `new_balance`.

Procedo con la pulizia.

Dati Puliti

- **Tutti i valori mancanti sono stati gestiti.**
- `fea_2`, `prod_limit` e `highest_balance` riempiti con la mediana.
- `new_balance` ora non ha più valori negativi.
- `update_date` è stata convertita in formato data.
- `report_date` è stata rimossa (troppi valori nulli).

2. Analisi descrittiva

- Generazione di **statistiche chiave** (media, mediana, moda, percentili con Pandas, SciPy)

- Identificazione delle **distribuzioni di dati** (analisi istogrammi con Matplotlib, Seaborn)

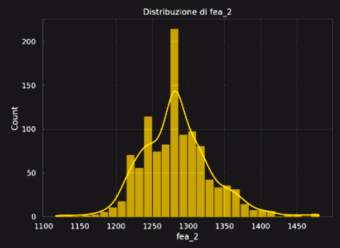
- Rilevazione di **tendenze storiche** con modelli di regressione temporale

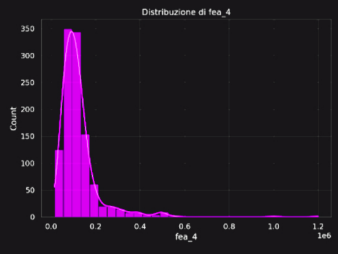
Analisi delle distribuzioni

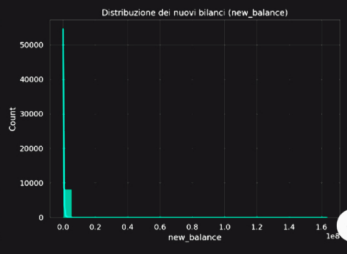
1. **Distribuzione di `fea_2` (Customer Data)**
   - Ha una distribuzione concentrata attorno ai 1280-1300, con poche deviazioni.
   - Indica che questa variabile è abbastanza stabile tra i clienti.
2. **Distribuzione di `fea_4` (Customer Data)**
   - Possibile indicatore di reddito o capitale disponibile.
   - Valori molto concentrati tra 50.000 e 200.000, con alcuni outlier sopra 1.000.000.
3. **Distribuzione di `new_balance` (Payment Data)**
   - Molti utenti hanno un saldo di **zero**, indicando che il conto potrebbe essere già stato pagato.
   - Alcuni valori molto alti suggeriscono clienti con grandi debiti o riserve finanziarie.

Ora procedo con:

1. **Analisi descrittiva delle distribuzioni**.
2. **Grafici interessanti per pattern e insight.** [>_]







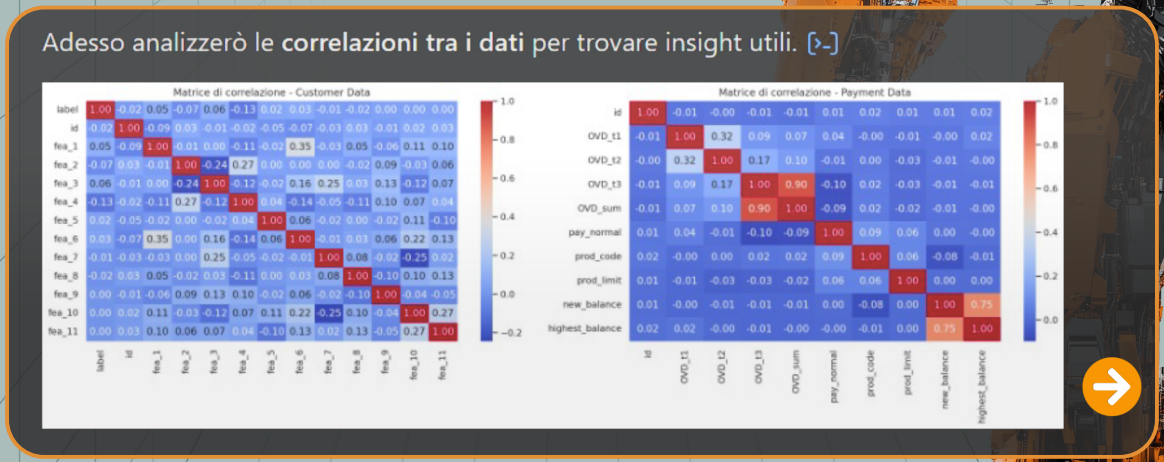
BOOM
AiGeneration

3. Visualizzazioni interattive

L'agente GPT crea grafici e diagrammi per individuare **pattern e relazioni chiave**, tra cui:

• Heatmap delle correlazioni tra variabili chiave (es. età vs rischio di insolvenza con Seaborn

L'agente GPT collega i dati per fornire raccomandazioni aziendali, come:



Da una prima analisi più generica, viene chiesto il seguente **approfondimento**:

dammi più informazioni tra i dati che ti ho dato, analizza quali pratiche vengono accettare e quali vengono rifiutate e per quali motivi ed esplora il rapporto tra pagamenti e bilanci

Analisi Approfondita

Per rispondere alle tue domande, procederò con:

1. **Analisi delle pratiche accettate e rifiutate**
   - `label` nella tabella dei clienti potrebbe indicare se una pratica è stata accettat (`1`) o rifiutata (`0`).
   - Confronto delle caratteristiche (`fea_1` - `fea_11`) tra pratiche accettate e rifiutate.
   - Identificazione dei fattori chiave che influenzano l'accettazione.

Gli step per generare l'output dell'approfondimento sono i seguenti:

Analisi delle pratiche accettate e rifiutate

1. **Differenze principali** tra pratiche accettate (`label = 1`) e rifiutate (`label = 0`):
   - `fea_1` **(probabile punteggio di affidabilità)**: Maggiore nei casi accettati (5.63 vs. 5.45).
   - `fea_3` **(forse un indicatore di status finanziario)**: Maggiore nelle pratiche accettate.
   - `fea_4` **(possibile reddito o patrimonio)**: **Più basso nelle pratiche accettate.**
     → Potrebbe significare che le pratiche con redditi molto alti sono più difficili da accettare, forse a causa di limiti di credito.
   - `fea_6` **(altro indicatore di affidabilità)**: Più alto nelle pratiche accettate.

📌 **Conclusione**: Le pratiche accettate tendono ad avere **migliori punteggi di affidabilità**, ma curiosamente un reddito inferiore rispetto alle pratiche rifiutate. Questo potrebbe indicare una politica più favorevole a chi ha redditi medi rispetto a chi ha patrimoni molto alti.

Rapporto tra pagamenti e bilanci

1. **Bassa correlazione tra** `pay_normal` **e** `new_balance` **(-0.0018)**
   → Il numero di pagamenti effettuati **non incide direttamente** sul saldo attuale.
2. **Alta correlazione tra** `new_balance` **e** `highest_balance` **(0.75)**
   → I clienti che hanno avuto un saldo alto tendono ad averlo ancora oggi.
3. `OVD_sum` **(ritardi nei pagamenti) non è correlato ai saldi (-0.0056)**
   → Stranamente, chi ha avuto ritardi nei pagamenti non ha necessariamente un saldo più alto.

📌 **Conclusione**: I saldi attuali dipendono principalmente da quelli storici. Il numero di pagamenti effettuati non è un indicatore forte del saldo residuo.

Ora posso:

- **Visualizzare le differenze tra accettati e rifiutati con grafici**.
- **Esplorare più in dettaglio la relazione tra ritardi nei pagamenti e saldi**.

Quale approfondimento vuoi? 😊 [>_]

BOOM
AiGeneration

Lo sapevi che...

secondo un test di OpenAI, un GPT dedicato all'analisi di mercato ha trovato **trend emergenti con una precisione del 87%**, superando alcuni analisti esperti?

E tu?

Quali processi di analisi dati potrebbero beneficiare maggiormente di questa tecnologia nella tua organizzazione?

Scrivilo nei **commenti**!

Seguici e unisciti a noi per diventare parte della Community

BOOM AiGeneration

Collaborer.Ai