Report Data Analysis

Analisi Studenti e Retention – Scuola di Musica

Clusterizzazione e Strategie di Re-Engagement

Sommario

[**INTRODUZIONE** 2](#_Toc202895661)

[**Pulizia dei Nominativi e Standardizzazione** 2](#_Toc202895662)

[**Integrazione con l'Anagrafica Studenti** 3](#_Toc202895663)

[**Gestione dei Nuclei Familiari** 3](#_Toc202895664)

[**Risultati e prossimi passi** 3](#_Toc202895665)

[**Pulizia e preparazione dei dati** 3](#_Toc202895666)

[**Creazione del Nucleo Familiare** 3](#_Toc202895667)

[**Passaggi in Power Query e Power BI** 5](#_Toc202895668)

## **INTRODUZIONE**

"*Un luogo dove la musica prende vita, tra le pareti di una seconda casa che accoglie, ispira e fa crescere".* L'analisi dati della Scuola di Musica[[1]](#footnote-1), è un progetto che unisce cuore e numeri per sostenere la transizione al nuovo gestionale (attivo da luglio 2024).

Utilizzando i dati storici anonimizzati (2019-2024) dei pagamenti di quote associative e corsi musicali - trasformati in insight attraverso ID univoci e chiavi primarie (PK) - il report si è focalizzato su tre pilastri tecnici:

**Andamento corsi e retention**

* *Tecniche applicate*: Analisi temporale con Power Query, calcolo metriche di fidelità (Churn Rate.
* *Risultato*: Mappatura dei tassi di abbandono e delle finestre critiche per interventi proattivi.

**Clusterizzazione avanzata**

* *Tecniche applicate*: implementato in Power BI di funzione per raggruppare studenti in base corso principale e anagrafica.
* *Risultato*: Identificazione di gruppi chiave.

**Ottimizzazione gestionale**

* *Tecniche applicate*: Data modeling per collegare dati storici (2019-2024) al nuovo sistema, creazione di colonne calcolate con DAX per colmare gap informativi.
* *Risultato*: Dashboard interattive per monitorare in tempo reale l’efficacia delle strategie.

## **Pulizia dei Nominativi e Standardizzazione**

Immagine che contiene testo, schermata, Carattere, numero

Il contenuto generato dall'IA potrebbe non essere corretto.Immagine che contiene testo, Carattere, linea, schermata

Il contenuto generato dall'IA potrebbe non essere corretto.Il progetto ha richiesto una fase iniziale di pulizia e preparazione dei dati, svolta principalmente in Python, per standardizzare i nominativi e garantire la qualità del dataset.

Nello specifico, dei fogli *quota\_associativa* e *freq\_corsi\_pagam*. Utilizzando Python e regex, sono stati identificati e corretti casi di nomi e cognomi attaccati (es. "MarcoRossi" → "Marco Rossi"), nonché valori con spazi eccessivi o formattazione incoerente. Un controllo rigoroso ha confermato l'assenza di errori nei nominativi del foglio *quota\_associativa*, mentre il foglio *freq\_corsi\_pagam* non presentava criticità. Sono stati inoltre individuati e gestiti nominativi con una sola parola (solo nome o solo cognome), potenzialmente problematici per l'analisi successiva.

## **Integrazione con l'Anagrafica Studenti**

Per collegare i pagamenti agli studenti registrati nel gestionale, è stato eseguito un merge tra i fogli *quota\_associativa*/*freq\_corsi\_pagam* e il file *Studenti\_rielaborato*, utilizzando l'*ID\_studente* come chiave. Questo ha permesso di verificare quanti record avevano un corrispondente nell'anagrafica:

- *239 valori univoci* in *quota\_associativa* e **288** in *freq\_corsi\_pagam* hanno evidenziato **38 nominativi distinti**, alcuni riconducibili a nuclei familiari (dove il pagamento era associato a un unico membro).

- I nominativi presenti solo in *freq\_corsi\_pagam* potrebbero indicare studenti che non hanno pagato la quota associativa separatamente, mentre quelli esclusivi di *quota\_associativa* potrebbero riflettere pagamenti accorpati con altri familiari.

Un grafico temporale mostrerà l'andamento dei pagamenti per identificare anomalie (es. quote non registrate) e trend annuali.

## **Gestione dei Nuclei Familiari**

L'analisi ha sfruttato la colonna *Indirizzo* per identificare i nuclei familiari, collegando studenti con lo stesso domicilio. Nei casi in cui il pagamento era associato a un genitore (es. mail con nome del figlio), i nominativi sono stati sostituiti per riflettere l'effettivo studente. Per gli studenti non presenti in *Studenti\_rielaborato*, è stato creato un *ID\_stud\_passati* per mantenere traccia anonima dei record storici. L'integrazione con il file *Nucleo\_familiare* ha permesso di arricchire il dataset con il numero di membri per famiglia (*Membri Familiari*), utile per analisi demografiche o promozioni group-based.

### **Risultati e prossimi passi**

Il processo ha garantito un dataset coerente, pronto per l'analisi delle dinamiche di pagamento e frequenza. Le criticità residue (es. 27 indirizzi mancanti) saranno affrontate con il cliente per migliorare la raccolta dati. Un'ulteriore ottimizzazione potrebbe includere:

- **Automazione delle sostituzioni** per ridurre l'intervento manuale nei futuri aggiornamenti.

## **Pulizia e preparazione dei dati**

Il progetto ha richiesto una fase iniziale di pulizia e preparazione dei dati, svolta principalmente in Python, per standardizzare i nominativi e garantire la qualità del dataset. Successivamente, il file *Studenti\_rielaborato* è stato elaborato in Power Query, dove sono state applicate diverse trasformazioni:

- La colonna *Data di nascita* è stata convertita in formato Data per facilitare il calcolo dell'età e l'utilizzo in formule condizionali.

- La colonna *Indirizzo* è stata resa univoca e utilizzata come chiave per identificare i membri di uno stesso nucleo familiare.

### **Creazione del Nucleo Familiare**

Per analizzare la composizione delle famiglie degli studenti, è stato creato un nuovo file *Nucleo*\_*familiare* raggruppando i dati per *Indirizzo* e *CAP*. Questo ha permesso di conteggiare il numero di membri per famiglia e assegnare un ID Indirizzo univoco come chiave primaria. Successivamente, è stato eseguito un merge tra *Studenti*\_*rielaborato* e *Nucleo*\_*familiare* per collegare ogni studente al proprio nucleo familiare. Dei 27 indirizzi mancanti, non è stato possibile recuperare informazioni complete, pertanto, non sono stati inclusi nel file finale.

**Integrazione dei Dati Storici**

Per garantire la completezza dell'anagrafica, sono stati integrati i dati storici degli studenti non più attivi. A partire dalle tabelle *quota* *associativa* e *pagamenti* *corsi*, sono stati eseguiti merge con il file *Studenti* per associare l'ID\_studente. I record senza *ID*\_*studente* sono stati filtrati, accodati e raggruppati per nome, assegnando loro un *ID*\_*stud*\_*passati* univoco. Questo identificatore è stato poi riportato nelle tabelle originali e utilizzato per creare una colonna personalizzata che unificasse gli ID attivi e passati. Infine, i file *Studenti*\_*rielaborato* e *studenti*\_*passati* sono stati mergiati per ottenere un'anagrafica completa, sebbene con alcune lacune informative.

Il processo di pulizia, trasformazione e integrazione dei dati ha permesso di ottenere un dataset coerente e pronto per l'analisi. L'utilizzo di chiavi univoche (*ID* *Indirizzo* e *ID\_studente*/*ID\_stud\_passati*) ha facilitato la correlazione tra le diverse tabelle, mentre l'attenzione ai dati mancanti ha evidenziato aree di miglioramento per la raccolta futura.

Sono state rielaborate tutte le tabelle, in cui sono state portate le PK e FK necessarie per le relazioni

### 

### **Passaggi in Power Query e Power BI**

**1. Unificazione dei Dati di Pagamento**

Per ottenere una visione completa delle transazioni, i due file dei pagamenti (*quota\_associativa* e *freq\_corsi\_pagam*) sono stati uniti in Power Query tramite **append**, creando un’unica tabella consolidata. È stata aggiunta una colonna **"Tipologia di pagamento"** per distinguere tra:

* **"Quota associativa"** (iscrizione alla scuola)
* **"Quota corsi"** (pagamento delle lezioni)

Questa operazione ha permesso di analizzare i flussi finanziari in modo più strutturato e di identificare eventuali discrepanze (es. studenti che hanno pagato solo la quota associativa senza corsi o viceversa).

**2. Pulizia e Arricchimento dell’Anagrafica Studenti e Corsi**

Sul file anagrafico sono state eseguite le seguenti operazioni:

* **Estrazione dell’anno di nascita**: La colonna "Data di nascita" è stata elaborata per ricavare solo l’anno, poi eliminata per semplificare il dataset.
* **Creazione della colonna "Età"**: Calcolata dinamicamente a partire dall’anno di nascita.
* **Definizione delle "Fasce d’età"**: Suddivise in categorie coerenti con il contesto della scuola (es. *Bambini 4-12*, *Teenager 13-17*, *Giovani Adulti 25-39*).

Queste modifiche hanno reso possibile analizzare la distribuzione degli studenti per età e adattare l’offerta formativa in base alle esigenze delle diverse fasce.

Nella tabella ‘Corsi’ per ottimizzare l’analisi dell’affluenza e la pianificazione dei corsi, sono stati suddivisi gli orari delle lezioni in **fasce temporali** (mattino, pomeriggio, sera), mantenendo sia la precisione degli intervalli (es. *10:00-12:00*) per i calcoli interni, sia etichette intuitive ("*Tarda mattinata*") per una lettura immediata, derivati dalla colonna [Orario] tramite condizioni in M Language.

**3. Risoluzione delle Relazioni Molti-a-Molti**

Per gestire correttamente le relazioni complesse tra:

* **Studenti ↔ Corsi** (uno studente può seguire più corsi, e un corso può avere più studenti)
* **Materie ↔ Docenti** (un insegnante può tenere più materie, e una materia può essere assegnata a più docenti)

Sono state create **tabelle bridge** in Power BI, che collegano le entità senza duplicare i dati. Questo approccio ha garantito un modello dati efficiente e privo di ambiguità.

**4. Creazione delle Tabelle dei Fatti e Calendari**

Sono state individuate tre tabelle principali dei fatti:

1. **Presenze** (frequenza alle lezioni)
2. **Ordini** (iscrizioni ai corsi)
3. **Pagamenti** (transazioni finanziarie)

Per ciascuna è stato generato un **calendario personalizzato**:

* **Anno scolastico (giugno-maggio)**: Poiché le iscrizioni da giugno si riferiscono all’anno successivo, è stata creata una colonna DAX per allineare i dati al calendario accademico.

**5. Sviluppo delle Misure per i KPI**

Sono state implementate misure DAX per monitorare:

* **Tasso di presenza** (presenze/assenze)
* **Retention rate** (studenti che rinnovano l’iscrizione)
* **Cluster CRM** (segmentazione studenti per fascia d’età e materia frequentata)

Queste metriche sono state visualizzate in dashboard interattive, con **slicer** per filtrare dinamicamente i dati in base a:

* Anno accademico
* Stato studente (attivo/non attivo)
* Tipologia di corso

**6. Visualizzazione dei Risultati**

I dati sono stati trasformati in:

* **Grafici temporali** (andamento iscrizioni e pagamenti)
* **Grafici a barre** (per materia e presenze)
* **Tabelle pivot** (analisi per cluster)

L’utilizzo di filtri interattivi permette di esplorare trend e anomalie, supportando decisioni basate sui dati per ottimizzare la gestione della scuola.

**Prossimi passi**: Approfondire l’analisi dei cluster per personalizzare le strategie di fidelizzazione.

1. *Nome inventato per privacy.* [↑](#footnote-ref-1)