

UNIVERSITÀ POLITECNICA DELLE MARCHE
FACOLTÀ DI INGEGNERIA
Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria Informatica e dell'Automazione



PROJECT MANAGEMENT

Progetto di Business Intelligence

Autore

Mattia Sbattella
Niccolò Ciotti
Luca Renzi

ANNO ACCADEMICO 2024-2025

Indice

1	Introduzione	1
1.1	Introduzione	1
1.2	Scelta del dataset	1
1.3	Preprocessing	2
1.3.1	Codice	3
1.4	Descrizione	5
2	Qlik	8
2.1	Introduzione	8
2.2	Come funziona	9
2.3	Foglio 1: Analisi sulla soddisfazione e sulla valutazione delle prestazioni	10
2.3.1	Analisi	11
2.4	Foglio 2: Relazione tra zona salariale e Work-Life Balance	12
2.4.1	Analisi	13
2.5	Relazione tra Performance Score, Satisfaction Score ed età per dipartimento	14
2.5.1	Analisi	15
2.6	Foglio 4: Analisi tra età e dipartimento	16
2.6.1	Analisi	17
3	Tableau	19
3.1	Introduzione	19
3.2	Come funziona	20
3.3	Dashboard 1: Variazione percentuale dei costi di training rispetto all'anno precedente	21
3.3.1	Analisi	22
3.3.2	Codice	24
3.4	Dashboard 2: Analisi sul costo medio della formazione	25
3.4.1	Analisi	26

3.5	Dashboard 3: Analisi del successo della formazione per fascia retributiva	28
3.5.1	Analisi	29
3.6	Dashboard 4: Analisi sul risultato della formazione	31
3.6.1	Analisi	33
4	Power BI	36
4.1	Introduzione	36
4.2	Come funziona	37
4.3	Foglio 1: Analisi delle assunzioni	38
4.3.1	Analisi	39
4.4	Foglio 2: Analisi delle dimissioni	42
4.4.1	Analisi	43
4.5	Foglio 3: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva	46
4.5.1	Analisi	47
4.6	Foglio 4: Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione	49
4.6.1	Analisi	50
5	Conclusioni	51
5.1	Conclusioni	51

Elenco delle figure

2.1	Analisi sulla soddisfazione e sulla valutazione delle prestazioni	10
2.2	Mappa che rappresenta la soddisfazione e la valutazione delle prestazioni	11
2.3	Analisi sulla soddisfazione e sulla valutazione delle prestazioni filtrata .	12
2.4	Mappa che rappresenta la soddisfazione e la valutazione delle prestazioni filtrata	12
2.5	Relazione tra zona salariale e Work-Life Balance	13
2.6	Relazione tra Performance Score, Satisfaction Score ed età per dipartimento	15
2.7	Relazione tra Performance Score, Satisfaction Score ed età per dipartimento con filtro	16
2.8	Analisi tra età e dipartimento	17
2.9	Analisi tra età e dipartimento con filtro	18
3.1	Variazione percentuale dei costi di training rispetto all'anno precedente	21
3.2	Variazione percentuale dei costi di training rispetto all'anno precedente filtrato	22
3.3	Modifica parametro data	24
3.4	Variazione percentuale dei costi di training rispetto all'anno precedente filtrato per anno	25
3.5	Analisi sul costo della formazione	26
3.6	Analisi sul costo della formazione filtrato per divisione	27
3.7	Analisi sul costo della formazione filtrato per anno e dipartimento . . .	28
3.8	Analisi del successo della formazione per fascia retributiva	29
3.9	Analisi del successo della formazione per fascia retributiva filtrato . . .	30
3.10	Analisi sul costo della formazione	32
3.11	Analisi sul risultato della formazione	33
3.12	Analisi costi della formazione	34
3.13	Analisi percentuali di successo della formazione	35

4.1	Distribuzione dei dipendenti per tipo di impiego e fascia retributiva.	38
4.2	Analisi delle assunzioni nel tempo, nel dipartimento Produzione per tipo di contratto.	39
4.3	Analisi delle assunzioni nel tempo, nel dipartimento Produzione per dipendenti a Contratto	40
4.4	Distribuzione dei dipendenti per tipo di impiego e fascia retributiva nell'anno 2022.	41
4.5	Analisi delle dimissioni.	42
4.6	Analisi delle dimissioni per dipartimento Sales.	43
4.7	Analisi delle dimissioni per durante gli anni con previsione futura	45
4.8	Distribuzione dei dipendenti per tipo di impiego e fascia retributiva.	46
4.9	Distribuzione dei dipendenti per tipo di impiego e fascia retributiva con filtro per etnia.	47
4.10	Distribuzione dei dipendenti per tipo di impiego e fascia retributiva in base allo stato.	48
4.11	Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione.	49
4.12	Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione per il dipartimento Software.	50

CAPITOLO 1

Introduzione

1.1 Introduzione

Nel contesto aziendale moderno, l'analisi dei dati delle Risorse Umane (HR) è fondamentale per prendere decisioni strategiche informate. Questo progetto di Business Intelligence si propone di sfruttare tre delle piattaforme più avanzate per la visualizzazione e l'analisi dei dati: Qlik, Tableau e Power BI.

L'obiettivo principale è trasformare i dati grezzi in insight significativi, nella previsione del turnover e nell'analisi delle prestazioni.

1.2 Scelta del dataset

Il dataset analizzato contiene informazioni dettagliate sui dipendenti di un'azienda, con particolare attenzione alle date di assunzione e uscita, la posizione lavorativa, l'unità di business, la soddisfazione, il bilanciamento tra lavoro e vita privata e le attività di formazione. La scelta del dataset è stata una fase cruciale dell'analisi, in quanto ha influenzato direttamente la qualità e la profondità degli insight ottenuti. Durante la selezione, sono stati considerati diversi fattori, tra cui la quantità di dati disponibili, la varietà delle informazioni contenute e la presenza di elementi temporali che permettessero di analizzare le tendenze nel tempo. Il primo dataset analizzato conteneva un numero molto elevato di righe, ma con pochi attributi. Questo limitava significativamente la capacità di effettuare analisi approfondite, poiché non era possibile identificare trend significativi. Inoltre, mancavano sequenze temporali, un elemento chiave delle nostre analisi.

Alla luce di queste considerazioni, si è optato per il dataset attuale, che, pur aven-

do un numero di righe inferiore, offriva una gamma molto più ampia di attributi e informazioni contestuali. Questo ha permesso di effettuare un'analisi più articolata e di esplorare con maggiore profondità i fenomeni aziendali come le assunzioni, la soddisfazione e la permanenza dei dipendenti in azienda.

L'importanza della combinazione tra quantità e qualità dei dati è risultata evidente nel corso del progetto: non sempre un dataset con molte righe è il più utile, soprattutto se privo di dettagli rilevanti per l'analisi. Nonostante ciò ci sono stati altri problemi legati al dataset emersi in corso d'opera, durante lo svolgimento delle varie analisi. Queste problematiche verranno approfondite nel capitolo 5.

1.3 Preprocessing

In fase di preprocessing, è stato necessario eliminare alcune colonne dal dataset originale per migliorare la qualità e la rilevanza dei dati. Le colonne eliminate sono le seguenti:

- *Unnamed: 0*: poiché rappresentava un semplice contatore per ogni dipendente e risultava poco rilevante, considerando già la presenza dell'attributo *EmployeeID*.
- *ADEmail*: essendo univoca per ciascun dipendente, l'email non apportava valore aggiunto per le analisi.
- *EmployeeClassificationType*: attributo ridondante, in quanto l'informazione era già contenuta nella colonna *EmployeeType*.
- *TerminationDescription*: attributo che rappresenta la motivazione della dimissione. Essendo una descrizione letterale e non categorizzata (esiste già *TerminationType*) è difficile fare delle analisi. Inoltre, molti dei valori sono nulli. Per queste motivazioni è stato scelto di eliminare l'attributo.

Successivamente, il dataset è stato sottoposto a una serie di trasformazioni per garantirne la coerenza e la leggibilità, utilizzando *Python* e la libreria *pandas* per manipolare i dati, insieme al modulo *datetime* per la gestione delle date. In questa fase si applicano trasformazioni utili per la pulizia e la standardizzazione dei dati. Si inizia con la definizione di un dizionario che associa le sigle degli stati americani ai rispettivi nomi completi, garantendo una maggiore leggibilità delle informazioni nel dataset. Successivamente, vengono definite tre funzioni fondamentali per la trasformazione dei dati. La prima funzione si occupa della conversione del formato della data di nascita, trasformandolo da "gg-mm-aaaa" a "gg-mese-aaaa", un formato più chiaro e coerente con le altre date nel dataset. La seconda funzione consente di sostituire le abbreviazioni degli stati con il nome completo, utilizzando il dizionario creato in precedenza, mentre la terza funzione calcola l'età dei dipendenti sulla base della data di nascita. Una volta definite le funzioni vengono applicate le trasformazioni, aggiornando le colonne *DOB*

per il formato della data di nascita, *State* per la conversione degli stati e *Age* per l'età calcolata. Il codice viene mostrato nella sezione 1.3.1.

1.3.1 Codice

```
1 import pandas as pd
2 from datetime import datetime
3
4 # Dizionario per la conversione delle sigle degli stati in nomi
5 # completi
6 state_abbreviations = {
7     "AL": "Alabama", "AK": "Alaska", "AZ": "Arizona", "AR": "
8         Arkansas", "CA": "California", "CO": "Colorado", "CT": "
9         Connecticut", "DE": "Delaware", "FL": "Florida", "GA": "
10        Georgia", "HI": "Hawaii", "ID": "Idaho", "IL": "Illinois",
11        ", "IN": "Indiana", "IA": "Iowa", "KS": "Kansas", "KY": "
12        Kentucky", "LA": "Louisiana", "ME": "Maine", "MD": "
13        Maryland", "MA": "Massachusetts", "MI": "Michigan", "MN": "
14        : "Minnesota", "MS": "Mississippi", "MO": "Missouri", "
15        "MT": "Montana", "NE": "Nebraska", "NV": "Nevada", "NH": "
16        "New_Hampshire", "NJ": "New_Jersey", "NM": "New_Mexico",
17        "NY": "New_York", "NC": "North_Carolina", "ND": "North_
18        Dakota", "OH": "Ohio", "OK": "Oklahoma", "OR": "Oregon",
19        "PA": "Pennsylvania", "RI": "Rhode_Island", "SC": "
20        South_Carolina", "SD": "South_Dakota", "TN": "Tennessee",
21        ", "TX": "Texas", "UT": "Utah", "VT": "Vermont", "VA": "
22        Virginia", "WA": "Washington", "WV": "West_Virginia", "
23        WI": "Wisconsin", "WY": "Wyoming"
24    }
25
26    def convert_dob_format(dob):
27        return datetime.strptime(dob, "%d-%m-%Y").strftime("%d-%b-%
28        Y")
29
30    def convert_state_abbreviation(state):
31        return state_abbreviations.get(state, state) # Ritorna il
32        nome completo se presente, altrimenti lascia invariato
33
34    def calculate_age(dob):
35        today = datetime.today()
36        birth_date = datetime.strptime(dob, "%d-%b-%Y")
```

```
18     age = today.year - birth_date.year - ((today.month, today.
19         day) < (birth_date.month, birth_date.day))
20
21 # Caricamento del dataset
22 df = pd.read_excel('/path/Messy_HR_Dataset_Detailed.xlsx')
23
24 # Conversione della colonna DOB
25 df["DOB"] = df["DOB"].apply(convert_dob_format)
26
27 # Conversione della colonna State
28 df["State"] = df["State"].apply(convert_state_abbreviation)
29
30 # Calcolo dell'eta e aggiunta della colonna Age
31 df["Age"] = df["DOB"].apply(calculate_age)
32
33 # Salvataggio del dataset modificato
34 df.to_excel('/path/HR_dataset.xlsx', index=False)
```

1.4 Descrizione

Di seguito sono riportati tutti gli attributi del dataset processato, il loro tipo, una piccola descrizione e alcuni dei valori assunti nel dataset.

Attributo	Tipo	Descrizione	Valori
FirstName	str	Nome del dipendente	-
LastName	str	Cognome del dipendente	-
StartDate	date	Data di inizio impiego	dd-month-yyyy (es. 16-Aug-22)
ExitDate	date	Data di uscita dall'azienda	dd-month-yyyy (es. 10-Nov-22) o EMPTY
Title	str	Titolo lavorativo del dipendente	Production Technician I, Production Technician II, Area Sales Manager, ...
Supervisor	str	Nome del supervisore	-
BusinessUnit	str	Unità di business	CCDR, EW, PL, TNS, ...
Employee Status	str	Stato del dipendente	Active, Future Start, Voluntarily Terminated
Employee Type	str	Tipo di impiego	Contract, Full-Time, Part-Time
PayZone	str	Zona salariale	Zone A, Zone B, Zone C,

Tabella 1.1: Descrizione dettagliata degli attributi del dataset

Attributo	Tipo	Descrizione	Valori
Termination Type	str	Tipo di terminazione	Unk, Involuntary, Voluntary, Resignation
Department Type	str	Dipartimento aziendale	Production, Sales, IT-IS ...
Division	str	Divisione aziendale	Finance, Accounting, Aerial, ...
DOB	date	Data di nascita	dd-month-yyyy (es. 30-Aug-1965)
State	str	Stato	-
JobFunction Description	str	Descrizione della funzione lavorativa	Accounting, Engineer, Assistant, Technician, ...
GenderCode	str	Genere	Female, Male
LocationCode	int	Codice della sede	-
RaceDesc	str	Origine etnica	White, Hispanic, Asian ...
MaritalDesc	str	Stato civile	Widowed, Single, Married, Divorced

Tabella 1.2: Descrizione dettagliata degli attributi del dataset

Attributo	Tipo	Descrizione	Valori
Performance Score	str	Valutazione delle performance	Fully Meets, Exceeds, Needs Improvement, PIP
Current Employee Rating	int	Punteggio del dipendente attuale	da 1 a 5
Employee ID	int	ID del dipendente	-
Survey Date	date	Data del sondaggio	dd-mm-yyyy (es. 09-09-2022)
Engagement Score	int	Punteggio di coinvolgimento	da 1 a 5
Satisfaction Score	int	Punteggio di soddisfazione	da 1 a 5
Work-Life Balance Score	int	Punteggio work-life balance	da 1 a 5
Training Date	date	Data della formazione	dd-month-yyyy (es. 30-Aug-1965)
Training Program Name	str	Nome del programma di formazione	Leadership Development, Customer Service, Project Management, ...
Training Type	str	Tipo di formazione	Internal, External
Training Outcome	str	Esito della formazione	Completed, Incomplete, Failed
Location	str	Località di lavoro	-
Trainer	str	Nome del formatore	-
Training Duration(Days)	int	Durata della formazione (giorni)	es. 10
Training Cost	float	Costo della formazione	es. 606.11

Tabella 1.3: Descrizione dettagliata degli attributi del dataset

CAPITOLO 2

Qlik



2.1 Introduzione

Qlik è una piattaforma di Business Intelligence che permette alle aziende di raccogliere, analizzare e visualizzare dati provenienti da fonti eterogenee, facilitando il processo decisionale grazie a una comprensione più profonda e immediata delle informazioni. Fondata su una tecnologia associativa unica, Qlik consente agli utenti di esplorare i dati in modo non lineare, identificando connessioni e tendenze che potrebbero rimanere nascoste con gli approcci tradizionali.

La piattaforma Qlik è composta principalmente da due soluzioni: *Qlik Sense* e *QlikView*. *Qlik Sense* è una moderna applicazione di visualizzazione dei dati che si distingue per la sua interattività e facilità d'uso, mentre *QlikView*, più orientato verso un'analisi predefinita, è spesso utilizzato per dashboard aziendali più strutturate e analisi ad hoc.

Uno degli aspetti chiave di Qlik è la sua capacità di integrarsi facilmente con una vasta gamma di fonti dati, dai database aziendali tradizionali a fonti cloud e social media, consentendo una visione completa e aggiornata del business. Grazie alla sua architettura associativa, Qlik offre agli utenti la possibilità di analizzare i dati da diverse prospettive, consentendo esplorazioni ad hoc e risposte rapide alle domande più complesse.

2.2 Come funziona

Qlik Sense è una piattaforma self-service che consente agli utenti di esplorare i dati e creare dashboard personalizzate senza bisogno di competenze avanzate di programmazione. Grazie alla sua interfaccia drag-and-drop, è possibile combinare diverse fonti di dati, generare analisi visive e ottenere insight in modo semplice e intuitivo.

Il primo passo è il caricamento dei dati, che può avvenire da diverse fonti e in vari formati, tra cui Excel, CSV o XML. Una volta importati, i dati vengono elaborati e associati automaticamente grazie al motore associativo di Qlik, che permette di esplorare le connessioni tra i dati senza vincoli gerarchici rigidi.

Qlik Sense offre una vasta gamma di oggetti per rappresentare i dati in modo chiaro e interattivo. Alcuni dei principali sono:

- *Grafici a barre*: ideali per confrontare valori tra diverse categorie.
- *Grafici a torta*: utilizzati per rappresentare la distribuzione percentuale di un insieme di dati.
- *Tabelle*: offrono una visione dettagliata dei dati grezzi con possibilità di filtri e ordinamenti.
- *Indicatori KPI*: mostrano metriche chiave con valori numerici e simboli visivi per evidenziare l'andamento.
- *Mappe geografiche*: utilizzate per visualizzare dati basati su coordinate geografiche, come vendite per regione.
- *Filtri e selettori*: consentono di interagire con i dati e restringere l'analisi su specifici parametri.
- *Alberi gerarchici*: utili per esplorare dati con una struttura a livelli, come organigrammi o relazioni tra prodotti.

Successivamente, questi oggetti vengono inseriti nei *fogli* di Qlik Sense, che rappresentano le pagine della dashboard. Ogni foglio può contenere diverse visualizzazioni e strumenti interattivi, consentendo di esplorare i dati da prospettive differenti e facilitando la suddivisione delle analisi in sezioni tematiche, permettendo di strutturare le informazioni in modo chiaro ed efficace.

In particolare, la nostra analisi contiene i seguenti fogli:

- Foglio 1: Analisi sulla soddisfazione e sulla valutazione delle prestazioni
- Foglio 2: Relazione tra zona salariale e Work-Life Balance
- Foglio 3: Relazione tra Performance Score, Satisfaction Score ed età per dipartimento
- Foglio 4: Analisi tra età e dipartimento

2.3 Foglio 1: Analisi sulla soddisfazione e sulla valutazione delle prestazioni

Come prima analisi viene descritta la dashboard sviluppata in Qlik Sense per l'analisi della soddisfazione e della valutazione delle prestazioni dei dipendenti. La dashboard illustrata in Figura 2.1 include due filtri principali:

- *EmployeeType*: consente di selezionare i dipendenti in base alla tipologia contrattuale, distinguendo, ad esempio, tra dipendenti a contratto, full-time e part-time.
- *DepartmentType*: permette di analizzare i dati filtrando per specifico dipartimento aziendale, consentendo confronti mirati tra le diverse aree dell'organizzazione.

Oltre ai filtri, la dashboard presenta due grafici a linee che illustrano l'andamento nel tempo di due indicatori chiave: *Satisfaction Score* ed *Employee Rating*.

Per fornire una sintesi immediata dei dati, sono stati inseriti due indicatori KPI:

- *Media del Satisfaction Score*: indica il valore medio della soddisfazione dei dipendenti per il contratto o il dipartimento selezionati.
- *Media dell'Employee Rating*: mostra la media della valutazione delle prestazioni per il contratto o il dipartimento selezionati.

Infine, la dashboard include una mappa geografica, Figura 2.2, che visualizza entrambi gli indicatori precedenti per ogni stato degli Stati Uniti. Attraverso una scala a gradiente è possibile rappresentare il livello di soddisfazione per ciascuno stato, mentre un sistema di punti permette di visualizzare il livello dell'Employee Rating.



Figura 2.1: Analisi sulla soddisfazione e sulla valutazione delle prestazioni

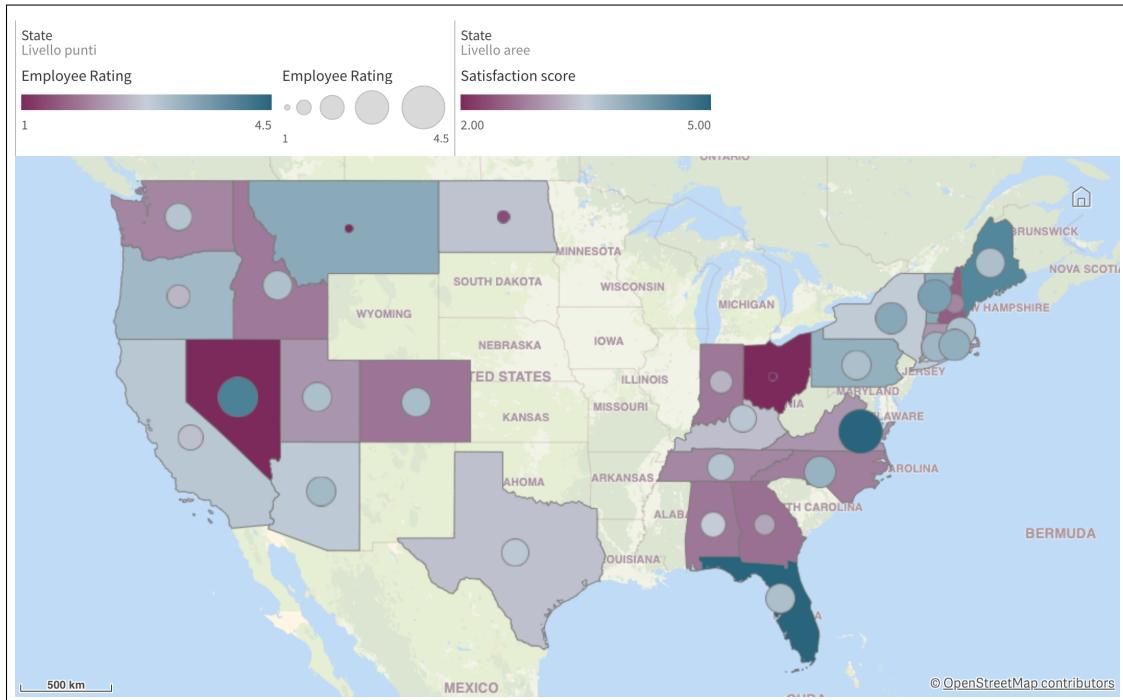


Figura 2.2: Mappa che rappresenta la soddisfazione e la valutazione delle prestazioni

2.3.1 Analisi

In particolare, si è scelto di analizzare la situazione dei dipendenti con contratto di tipo *Part-Time* appartenenti al dipartimento *IT/IS*, come mostrato in Figura 2.3. Applicando questi filtri nella dashboard, è possibile osservare l'andamento nel tempo dei due indicatori principali: il *Satisfaction Score* e l'*Employee Rating*, entrambi caratterizzati da un andamento discontinuo, con variazioni periodiche che evidenziano fluttuazioni nei livelli di soddisfazione e nelle prestazioni.

Nonostante queste fluttuazioni, grazie ai KPI associati a questa selezione, è possibile osservare i valori medi dei due indicatori: il *Satisfaction Score medio*, pari a 2.87, indica un discreto livello di soddisfazione, mentre l'*Employee Rating medio*, pari a 2.94, evidenzia anch'esso una performance soddisfacente per il dipartimento *IT/IS* nei contratti *Part-Time*.

La mappa geografica, aggiornata in base ai filtri selezionati, consente di esaminare la distribuzione della soddisfazione e della valutazione delle prestazioni per stato. Grazie alla scala a gradiente, si può osservare che in Oregon il livello di soddisfazione è circa 2,5, così come quello delle prestazioni. In Texas, entrambi i valori si aggirano intorno al 3, così come in Connecticut e Massachusetts. Infine, in Kentucky, il valore della soddisfazione è circa 2,5 e quello delle prestazioni circa 3.

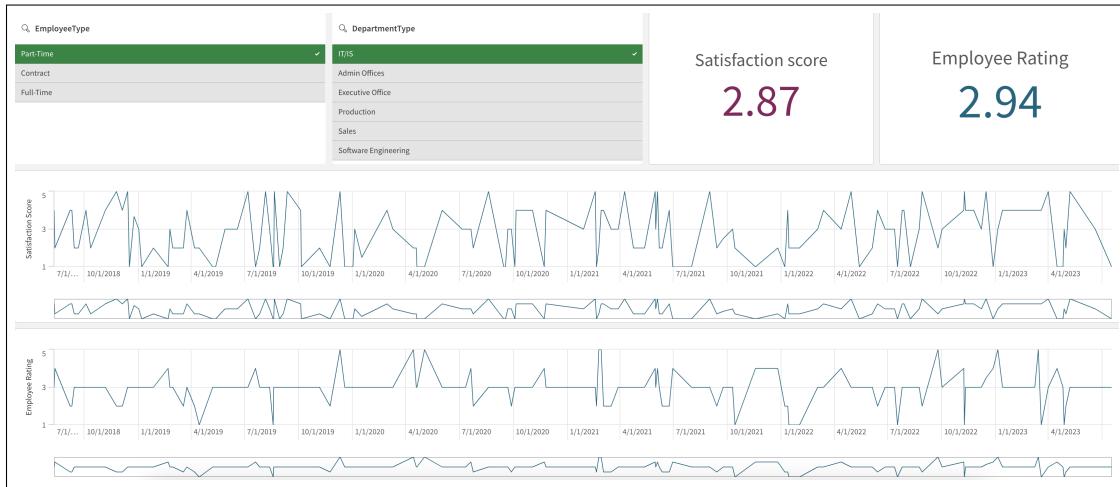


Figura 2.3: Analisi sulla soddisfazione e sulla valutazione delle prestazioni filtrata

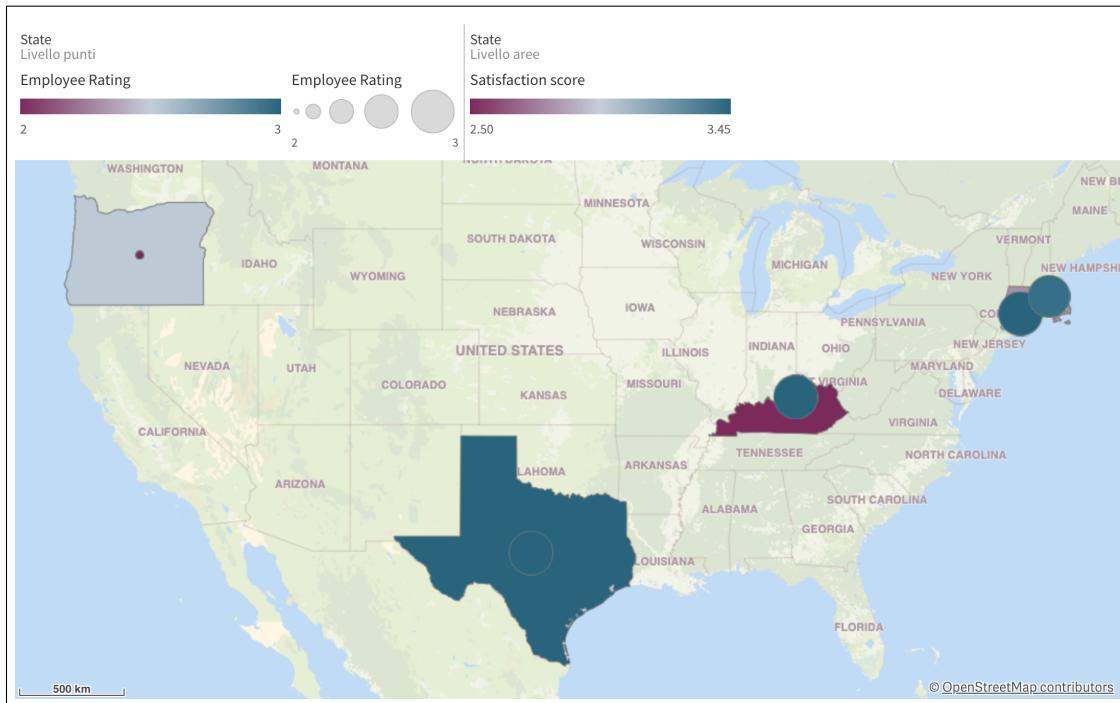


Figura 2.4: Mappa che rappresenta la soddisfazione e la valutazione delle prestazioni filtrata

2.4 Foglio 2: Relazione tra zona salariale e Work-Life Balance

La dashboard illustrata in Figura 2.5 è progettata per esplorare la connessione tra la zona salariale (Pay Zone) e il bilanciamento tra vita lavorativa e vita privata dei

dipendenti (rappresentato da valori compresi tra 1 e 5).

Il foglio include un filtro per *Pay Zone*, che consente di selezionare specifiche fasce salariali e visualizzare i dati relativi esclusivamente ai dipendenti appartenenti alla zona salariale selezionata. In questo modo, è possibile analizzare come il livello retributivo influisca sul *Work-Life Balance*.

Inoltre, un grafico a torta mostra la distribuzione del *Work-Life Balance* tra tutti i dipendenti, evidenziando le diverse categorie di bilanciamento tra lavoro e vita privata. Ciò consente di comprendere come la percezione di equilibrio sia distribuita all'interno dell'azienda, anche grazie all'inserimento di un KPI che visualizza la media del *Work-Life Balance* per l'intero set di dati.

Infine, un diagramma a barre raggruppato presenta, lungo l'asse delle ascisse, i due indicatori principali: *Soddisfazione* e *Work-Life Balance*, mentre l'asse delle ordinate mostra le diverse divisioni aziendali. Questo grafico consente di confrontare il livello di soddisfazione e il bilanciamento tra vita lavorativa e privata per ciascuna divisione, fornendo una panoramica chiara e immediata delle differenze tra i vari dipartimenti.

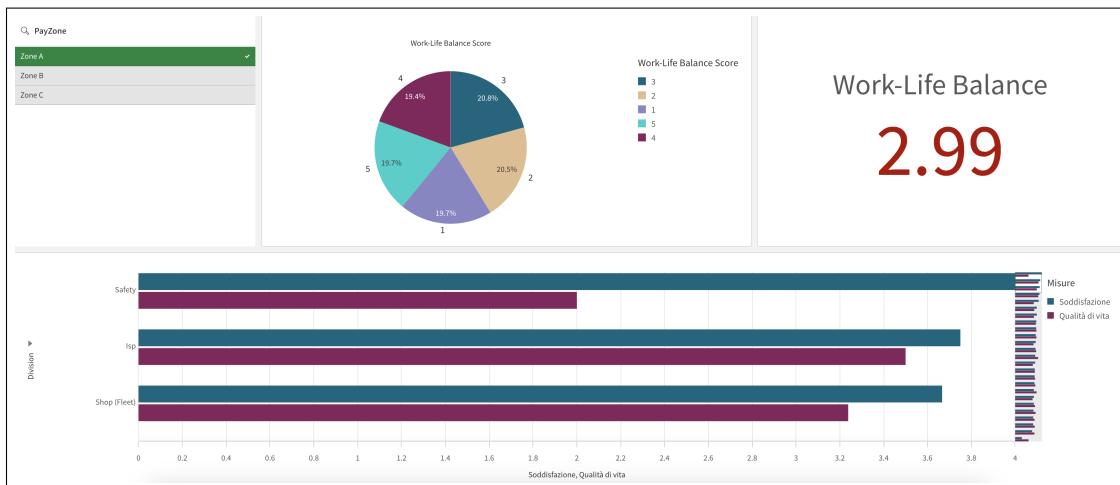


Figura 2.5: Relazione tra zona salariale e Work-Life Balance

2.4.1 Analisi

Nel caso specifico, applicando un filtro sulla zona salariale A, la più alta, si osserva che circa il 20% dei dipendenti presenta un bilanciamento tra vita lavorativa e privata distribuito tra le diverse classi del *Work-Life Balance Score*. Nonostante la retribuzione elevata, il valore medio di *Work-Life Balance* si attesta a 2.99, indicando un livello discreto di equilibrio tra lavoro e vita personale.

Dal grafico sottostante emerge che le divisioni con i punteggi più alti sia in termini di soddisfazione sia di *Work-Life Balance* sono *Safety*, *ISP* e *Shop*. Tuttavia, è interessante notare che, sebbene *Safety* registri il valore più elevato di soddisfazione, il suo *Work-Life*

Balance è il più basso tra le tre divisioni. In generale, per tutte e tre le aree analizzate, la soddisfazione lavorativa risulta sempre superiore al bilanciamento tra vita e lavoro.

Questa analisi suggerisce che un salario più alto non garantisce necessariamente un migliore equilibrio tra vita professionale e personale. Il punteggio medio di *Work-Life Balance*, pari a 2.99, potrebbe essere influenzato da carichi di lavoro più elevati. Inoltre, la discrepanza tra soddisfazione e *Work-Life Balance* nelle divisioni *Safety*, *ISP* e *Shop* evidenzia come, pur essendo appagati dal proprio impiego, i dipendenti percepiscano una gestione meno ottimale del tempo libero. In particolare, la divisione *Safety*, che mostra il massimo livello di soddisfazione e il minimo di *Work-Life Balance*, potrebbe riflettere un'intensità lavorativa significativamente superiore alle altre.

2.5 Relazione tra Performance Score, Satisfaction Score ed età per dipartimento

Il terzo foglio della dashboard, Figura 2.6, è progettato per analizzare la distribuzione e l'interazione di questi indicatori all'interno dell'organizzazione.

La visualizzazione è organizzata attraverso un filtro gerarchico che consente di selezionare i dati in base a *DepartmentType*, *Division* e *Job Function*, permettendo un'analisi mirata su specifici gruppi di dipendenti.

Nella parte centrale è presente un diagramma a barre che mostra la relazione tra *Satisfaction Score* e *Performance Score*, con il primo rappresentato sull'asse delle ascisse e il secondo sulle ordinate. I livelli di *Performance Score* considerati sono quattro:

- *Fully Meets*: Il dipendente soddisfa pienamente le aspettative del ruolo, raggiungendo gli obiettivi richiesti.
- *Exceeds*: Il dipendente supera le aspettative, dimostrando prestazioni eccellenti.
- *Needs Improvement*: Il dipendente non raggiunge del tutto gli standard richiesti e necessita di miglioramenti.
- *PIP* (Performance Improvement Plan): Il dipendente ha prestazioni insufficienti ed è inserito in un piano di miglioramento per mantenere il posto di lavoro.

Questo diagramma consente di individuare come la soddisfazione dei dipendenti si distribuisce in funzione delle loro prestazioni.

Inoltre, è stata inserita una mappa ad albero che rappresenta il numero di dipendenti per ciascuna categoria di *Performance Score*, offrendo una visione immediata della distribuzione del personale in base alla valutazione delle performance.

Infine, nella parte inferiore possiamo osservare un grafico a dispersione che mostra la relazione tra età e *Performance Score*, visualizzando per ciascuna categoria la media dell'età dei dipendenti. Questo permette di analizzare eventuali tendenze legate all'esperienza e alla valutazione delle performance.

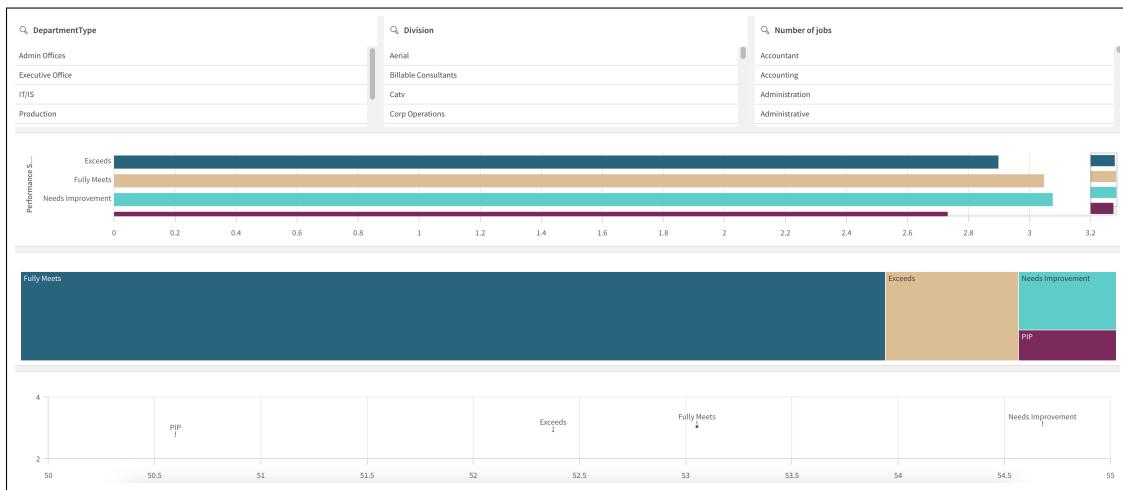


Figura 2.6: Relazione tra Performance Score, Satisfaction Score ed età per dipartimento

2.5.1 Analisi

In questa analisi, Figura 2.7, sono stati applicati i filtri *Department Type: Admin Offices* e *Division: Executive e Financial & Accounting* per esaminare la relazione tra *Performance Score*, *Satisfaction Score* ed età in questo specifico contesto organizzativo.

Il diagramma a barre mostra che, all'interno del sottoinsieme filtrato, sono presenti solo due categorie di *Performance Score*, suggerendo che le prestazioni dei lavoratori in queste divisioni sono generalmente positive:

- Exceeds, con un *Satisfaction Score* pari a 2
- Fully Meets, con un *Satisfaction Score* di 3.75

La differenza tra il *Satisfaction Score* di 2 per chi rientra nella categoria *Exceeds* e il valore di 3.75 per quelli di *Fully Meets* potrebbe riflettere una diversa percezione della soddisfazione lavorativa, con punteggi più alti associati a una performance che soddisfa pienamente le aspettative.

La mappa ad albero evidenzia la distribuzione dei dipendenti tra queste categorie: quattro dipendenti rientrano nella valutazione *Fully Meets*, mentre un solo dipendente ha ottenuto un punteggio *Exceeds*.

Infine, il grafico a dispersione mostra l'età media per ciascuna categoria di *Performance Score*. Si osserva che i dipendenti con valutazione *Fully Meets* hanno un'età media di circa 40 anni, mentre coloro che rientrano nella categoria *Exceeds* hanno un'età media di 52 anni. Questa distribuzione suggerisce che, in questo specifico gruppo, i dipendenti con valutazioni di performance più elevate tendono ad avere un'età media superiore, forse grazie all'esperienza accumulata nel tempo. Questo potrebbe indicare che l'esperienza ha un impatto positivo sulla valutazione delle prestazioni in queste divisioni specifiche, con una maggiore soddisfazione associata ai dipendenti più esperti.

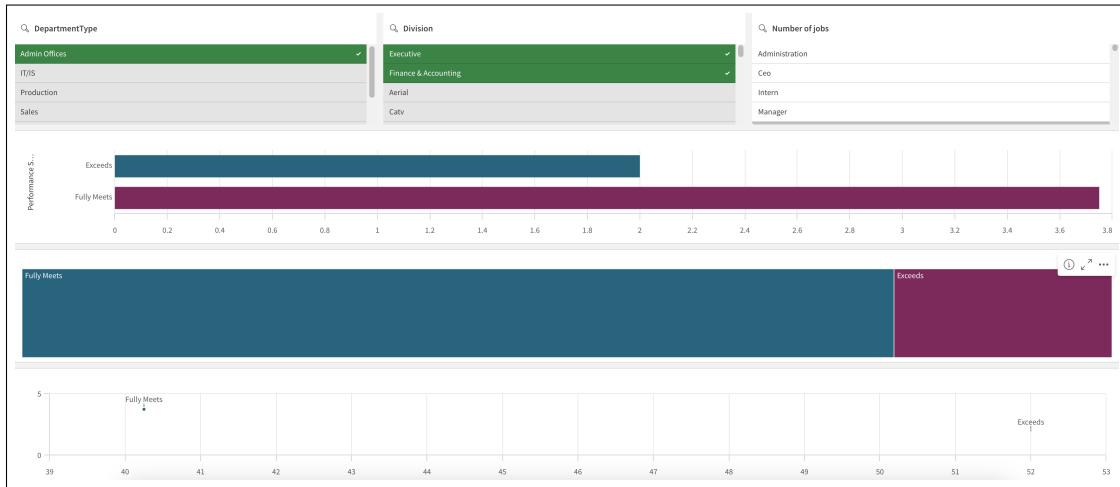


Figura 2.7: Relazione tra Performance Score, Satisfaction Score ed età per dipartimento con filtro

2.6 Foglio 4: Analisi tra età e dipartimento

La quarta dashboard, Figura 2.8, è dedicato all’analisi della relazione tra l’età dei dipendenti e il loro dipartimento di appartenenza. Anche in questo caso, attraverso il filtro gerarchico *DepartmentType*, *Division* e *Job Function*, è possibile selezionare specifici gruppi organizzativi per una valutazione più mirata.

La dashboard presenta un grafico a barre che mostra la media dell’età per ciascuna divisione, consentendo di confrontare le differenze tra le varie aree aziendali. Accanto a questo, un grafico a torta rappresenta la distribuzione dell’età in base alla *Job Function*, offrendo una visione sulla ripartizione dell’età nei diversi ruoli aziendali. Infine, nella parte inferiore del foglio, un grafico a dispersione evidenzia il rapporto tra il numero di dipendenti per ogni *Department* e la relativa età media, con l’asse delle ordinate che rappresenta il numero di lavoratori e quello delle ascisse l’età media.

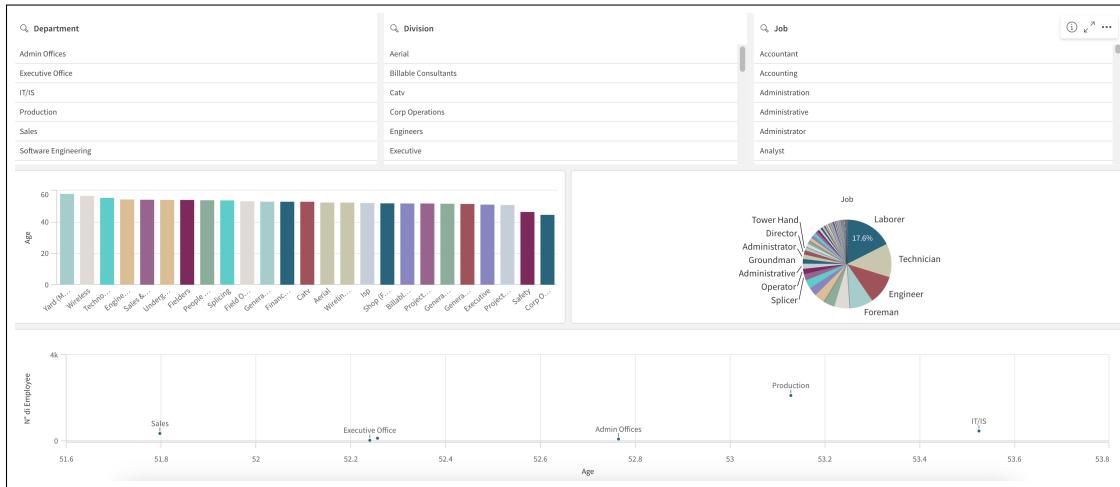


Figura 2.8: Analisi tra età e dipartimento

2.6.1 Analisi

In questo caso specifico, Figura 2.9, l'analisi è stata effettuata selezionando i dipartimenti *IT/IS* e *Production*, le divisioni *Engineers*, *Project Management* e *Technology/IT*, e le *Job Functions* *Engineer*, *Program Manager* e *Project Manager*.

Il grafico a barre evidenzia che l'età media dei dipendenti nelle divisioni *Technology/IT* e *Engineers* si aggira intorno ai 50 anni, mentre nella divisione *Project Management* è significativamente più bassa, circa 35 anni. Il grafico a torta, invece, fornisce una panoramica sulla distribuzione dell'età per funzione lavorativa: il 20,4% dei *Program Manager* ha un'età media di circa 50 anni, il 75,8% degli *Engineer* si attesta intorno ai 53 anni, mentre solo il 3,8% dei *Program Manager* ha un'età media di 36 anni. Infine, nel grafico a dispersione, si osserva che circa 150 dipendenti del dipartimento *Production* hanno un'età media di 50 anni, mentre 40 dipendenti della sezione *IT/IS* raggiungono un'età media di circa 56 anni.

Questa analisi mette in evidenza alcune tendenze interessanti all'interno dell'organizzazione. In particolare, si nota una concentrazione di dipendenti più anziani nei ruoli tecnici e ingegneristici, suggerendo che queste posizioni richiedano una maggiore esperienza e una permanenza più lunga all'interno dell'azienda. Al contrario, la divisione *Project Management* presenta un'età media inferiore, il che potrebbe indicare una maggiore mobilità o un accesso più rapido a ruoli di responsabilità per i dipendenti più giovani. Inoltre, la differenza di età tra i dipartimenti *Production* e *IT/IS* suggerisce che in ambito produttivo si possa avere una maggiore stabilità occupazionale nel tempo, mentre nei ruoli IT il personale tende a essere leggermente più anziano, probabilmente a causa delle competenze richieste e dell'esperienza necessaria per ricoprire determinate funzioni.

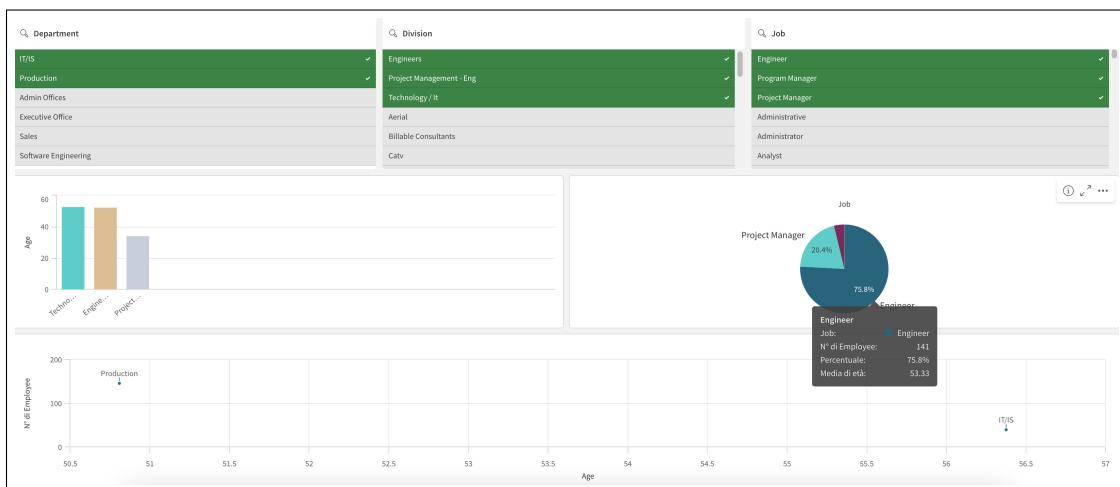


Figura 2.9: Analisi tra età e dipartimento con filtro

CAPITOLO 3

Tableau



3.1 Introduzione

Tableau è una piattaforma di Business Intelligence incentrata sulla parte di data visualization dando la possibilità di per trasformare dati complessi in dashboard intuitive e interattive. Tableau è uno degli strumenti più diffusi nel settore dell'analisi dei dati grazie alla sua capacità di trasformare dataset complessi in dashboard interattive e report di facile interpretazione, ampiamente utilizzato nel mondo aziendale per il monitoraggio delle performance, la reportistica avanzata e l'analisi predittiva.

Rispetto ad altri strumenti come Qlik, Tableau, attraverso un interfaccia intuitiva e drag-and-drop, si distingue per la sua semplicità nella creazione di dashboard personalizzate che facilitano una maggiore comprensione dei dati attraverso un'ampia gamma di grafici, mappe e tabelle. Inoltre, integrando strumenti di intelligenza artificiale e analisi avanzata consente la generazione di statistiche avanzate per individuare tendenze e anomalie dei dati. Grazie alla facilità di personalizzazione ci permette di ottenere analisi predittive utilizzando strumenti di analisi avanzata per identificare tendenze, prevedere ritardi e mitigare i rischi con modelli di previsione basati sui dati.

3.2 Come funziona

Tableau è una piattaforma di Business Intelligence incentrata sulla parte di data visualization dando la possibilità di per trasformare dati complessi in dashboard intuitive e interattive in modo semplice e strutturato.

Il primo passo è il caricamento dei dati, che può avvenire in diverse fonti e in vari formati tra cui file Excel, file di testo, file JSON, file PDF, fornisce anche la possibilità di importare e collegare tra loro dataset eterogenei, database e altre fonti. Tableau offrono due modalità di connessione ai dati:

- *Live*: è preferibile quando la sorgente dati è veloce ed ottimizzata per analisi interattive, o quando è necessario visualizzare dati aggiornati in tempo reale in dataset con grandi mole di dati.
- *In-Memory*: ideale quando la sorgente dati è troppo lenta per analisi interattive o quando bisogna lavorare offline con i dati salvati. Tableau utilizza le risorse hardware disponibili, come RAM e disco, per ottimizzare le prestazioni e consentire l'elaborazione efficiente di dataset di grandi dimensioni.

Analogamente a quanto fatto con Qlik, il dataset è stato importato in Tableau con la modalità In-Memory. Tableau offre una vasta gamma di oggetti per rappresentare i dati in modo chiaro e interattivo. Alcuni dei principali sono:

- *Grafici a barre*: ideali per confrontare valori tra diverse categorie.
- *Grafici a torta*: utilizzati per rappresentare la distribuzione percentuale di un insieme di dati.
- *Tabelle*: offrono una visione dettagliata dei dati grezzi con possibilità di filtri e ordinamenti.
- *Indicatori KPI*: mostrano metriche chiave con valori numerici e simboli visivi per evidenziare l'andamento.
- *Mappe geografiche*: utilizzate per visualizzare dati basati su coordinate geografiche, come vendite per regione.
- *Filtri e selettori*: consentono di interagire con i dati e restringere l'analisi su specifici parametri.
- *Alberi gerarchici*: utili per esplorare dati con una struttura a livelli, come organigrammi o relazioni tra prodotti.

Successivamente, questi oggetti vengono inseriti nei fogli di lavoro (Worksheet) di Tableau che compongono la dashboard. Ogni foglio può contenere diverse visualizzazioni e strumenti interattivi, consentendo l'esplorazione i dati da differenti prospettive, permettendo la suddivisione delle analisi in sezioni tematiche, raggruppando le informazioni in modo chiaro ed efficace.

In particolare, la nostra analisi contiene le seguenti dashboard:

- Dashboard 1: Variazione percentuale dei costi di training rispetto all’anno precedente
- Dashboard 2: Analisi sul costo della formazione
- Dashboard 3: Analisi del successo della formazione per fascia retributiva
- Dashboard 4: Analisi sul risultato della formazione

3.3 Dashboard 1: Variazione percentuale dei costi di training rispetto all’anno precedente

Come prima analisi viene descritta la dashboard sviluppata in Tableau per l’analisi della variazione percentuale dei costi di training rispetto all’anno precedente. L’obiettivo è comprendere come gli investimenti nella formazione siano cambiati nel tempo, valutando eventuali aumenti, riduzioni o stabilità nelle diverse aree dell’azienda, aiutando a individuare trend di crescita, stabilità o riduzione delle spese. Nella dashboard illustrata in Figura 3.1 sono presenti diversi filtri:

- *Anno*: permette di analizzare i dati filtrando per lo specifico anno.
- *Departement Type*: permette di analizzare i dati filtrando per specifico dipartimento aziendale, consentendo confronti mirati tra le diverse aree dell’organizzazione.
- *Division*: consente di selezionare specifiche divisioni aziendali all’interno dei dipartimenti.

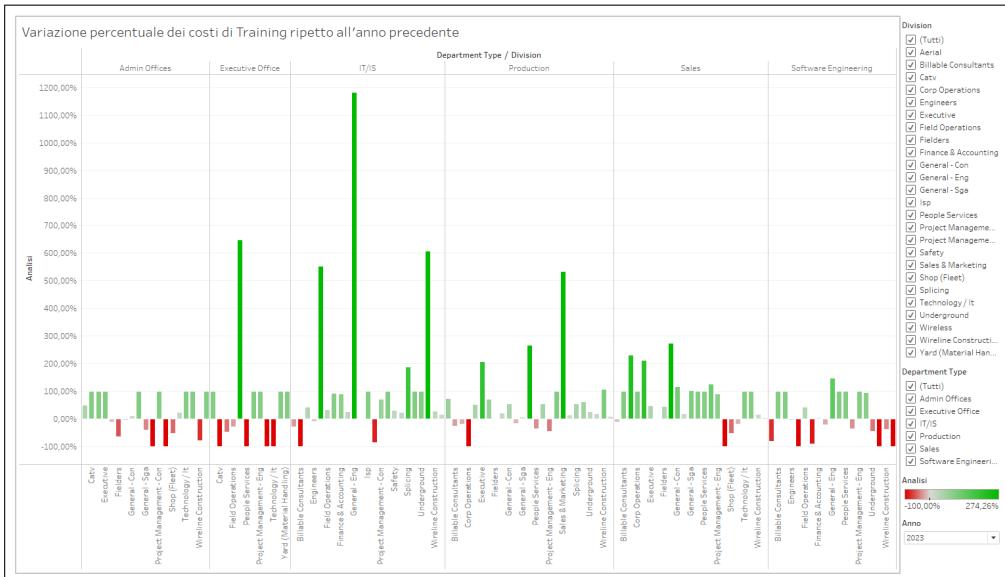


Figura 3.1: Variazione percentuale dei costi di training rispetto all’anno precedente

3.3.1 Analisi

In particolare, si è scelto di studiare i costi di formazione in base alla tipologia del dipartimento *Admin*, *IT/IS* e *Software Engineer* e le loro rispettive divisioni aziendali. Questo implica che l’analisi è focalizzata sull’area tecnologica e sulla gestione esecutiva, escludendo altre sezioni aziendali come vendite, produzione e amministrazione.

Dai dati di Figura 3.2 emerge che nell’anno 2023 si ha una significativa variabilità nei costi di formazione, con alcune divisioni che hanno aumentato il budget in modo significativo, superiori al 500%, e altre che lo hanno ridotto drasticamente, fino al -100%. Il budget di formazione non è stato aumentato in modo uniforme, ma piuttosto in base a necessità specifiche dei dipartimenti.

La concentrazione delle risorse in dipartimenti, come *IT/IS* ed *Executive Office*, potrebbero indicare una priorità strategica nell’innovazione tecnologica e nella leadership aziendale, mentre gli azzeramenti dei costi di formazione in alcune aree, espresse con barre rosse, potrebbe essere dovuto a una riorganizzazione interna o alla decisione di ridistribuire il budget verso altre priorità.

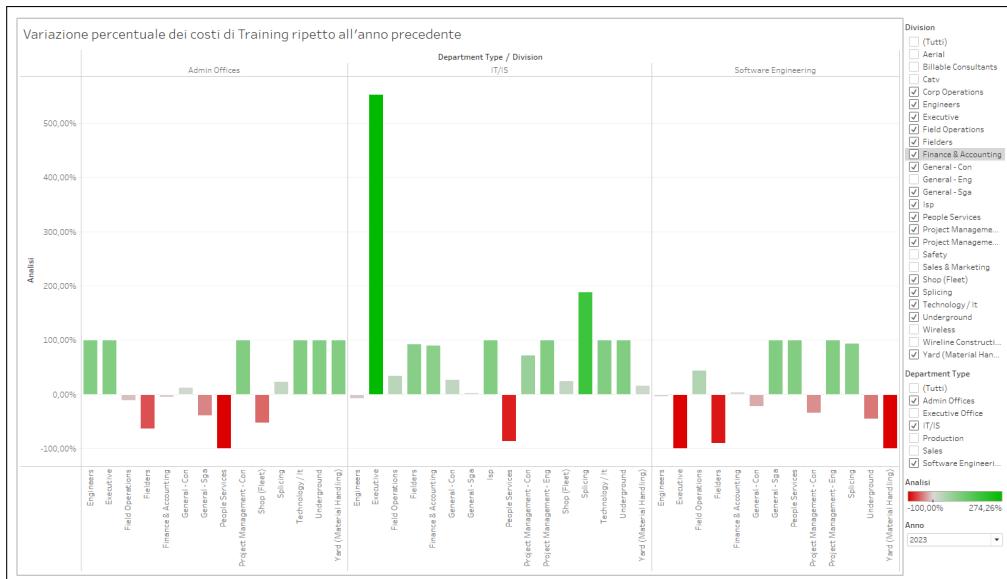


Figura 3.2: Variazione percentuale dei costi di training rispetto all’anno precedente filtrato

Invece nella Figura 3.4 abbiamo analizzato la variazione dei costi di formazione in base all’anno. Selezionando l’anno 2022 nella scheda rappresentata in Figura 3.3, confrontiamo i dati con quelli dell’anno precedente, il 2021. Questa schermata consente di configurare il parametro “Anno”, utilizzato per filtrare un’analisi in base all’anno selezionato. Il parametro è di tipo *Data* i valori possono essere visualizzati nel formato “gg/mm/aaaa” o in uno dei formati di visualizzazione predefiniti. L’utente può scegliere un valore dall’elenco predefinito, che include le date di riferimento per gli

anni disponibili, compresi tra 2021 e 2024. Inoltre, è possibile selezionare la modalità di visualizzazione dei dati tra tre opzioni:

- *Elenco*, che permette di scegliere tra anni specifici.
- *Intervallo*, che consente di impostare un range di date.
- *Tutto*, per visualizzare tutti i dati senza restrizioni.

Questa flessibilità permette di adattare il filtro alle esigenze dell’analisi, facilitando l’esplorazione e il confronto dei dati nel tempo.

Dal report emerge che, nella maggior parte dei casi, il valore della variazione percentuale dei costi di formazione è pari al 100%. Questo è dovuto dalla scarsità di dati disponibili per il 2021, suggerendo che gli investimenti nella formazione siano stati effettuati principalmente nel 2022 e in anni successivi. Nelle divisioni in cui non compare alcuna barra, la variazione è pari allo 0%, indicando l’assenza di differenze rispetto all’anno precedente.

Selezionando l’anno 2023, notiamo una maggiore quantità di dati e una distribuzione degli investimenti meno uniforme, fornendo un quadro più dettagliato e comprensibile. Di seguito troviamo nella sezione 3.3.2 il codice utilizzato per poter effettuare questa analisi.

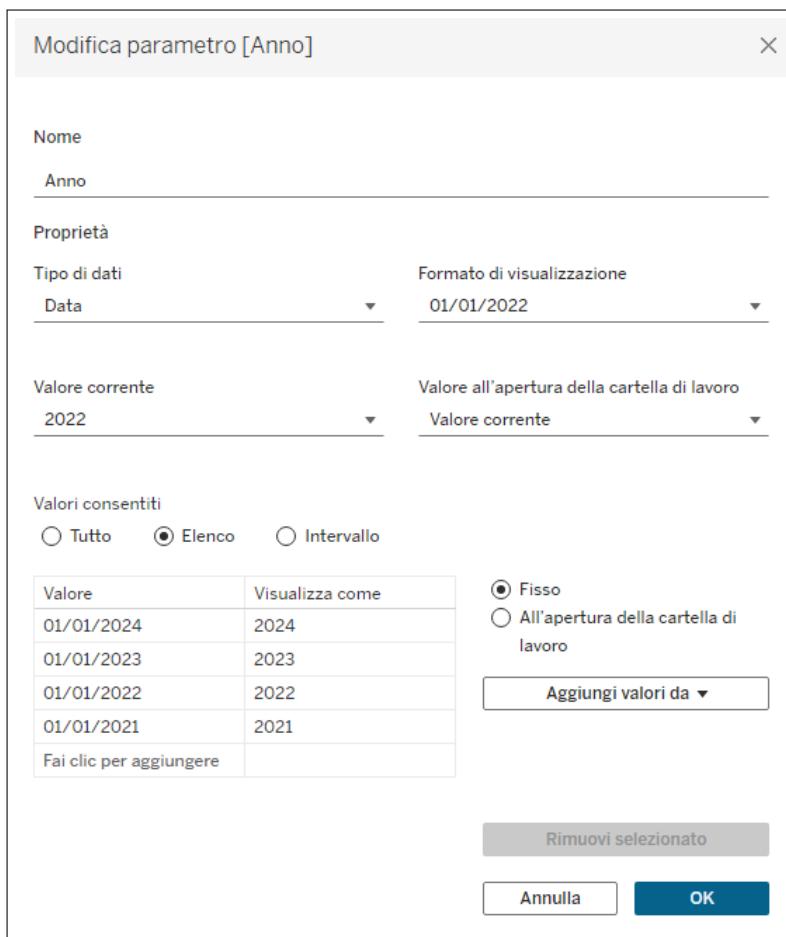


Figura 3.3: Modifica parametro data

3.3.2 Codice

Se non ci sono dati per entrambi gli anni, mostra 0%

```

1 IF ZN(SUM(IF YEAR([Training Date]) = YEAR([Anno]) THEN [
    Training Cost] END)) = 0 AND ZN(SUM(IF YEAR([Training Date]
    ]) = YEAR([Anno]) - 1 THEN [Training Cost] END)) = 0
2 THEN 0

```

Se l'anno attuale ha dati ma l'anno precedente no, mostra 100%

```

1 ELSEIF ZN(SUM(IF YEAR([Training Date]) = YEAR([Anno]) THEN [
    Training Cost] END)) > 0 AND ZN(SUM(IF YEAR([Training Date]
    ]) = YEAR([Anno]) - 1 THEN [Training Cost] END)) = 0
2 THEN 1

```

Se l'anno attuale non ha dati ma l'anno precedente sì, mostra -100%

```

1 ELSEIF ZN (SUM(IF YEAR([Training Date]) = YEAR([Anno]) THEN (
    Training Cost) END)) = 0 AND ZN (SUM (IF YEAR ([Training
    Date]) = YEAR([Anno]) - 1 THEN (Training Cost] END)) > 0
2 THEN (0 - ZN(SUM(IF YEAR ([Training Date]) = YEAR([Anno]) - 1
    THEN [Training Cost] END))) / ZN(SUM(IF YEAR([Training Date]
    )
3 = YEAR ([Anno]) - 1 THEN [Training Cost] END))

```

Altrimenti calcola normalmente la variazione percentuale

```

1 ELSE
2 (ZN (SUM (IF YEAR ([Training Date]) = YEAR([Anno]) THEN [
    Training Cost] END)) - ZN(SUM(IF YEAR ([Training Date]) =
    YEAR ([Anno]) - 1
3 THEN (Training Cost] END))) / ZN(SUM(IF YEAR ([Training Date])
    = YEAR([Anno]) - 1 THEN [Training Cost] END) )
4 END

```



Figura 3.4: Variazione percentuale dei costi di training rispetto all'anno precedente filtrato per anno

3.4 Dashboard 2: Analisi sul costo medio della formazione

In questo foglio della nostra analisi in Tableau si confrontano il costo medio dei vari dipartimenti aziendali e le relative figure lavorative, fornendo la possibilità di scegliere la tipologia della formazione e l'intervallo di tempo, espresso in trimestri. Nella dashboard in Figura 3.5. Nell'analisi sono presenti diversi filtri:

- *Division*: consente di selezionare specifiche divisioni aziendali all'interno dei dipartimenti.
- *Department Type*: permette di analizzare i dati filtrando per specifico dipartimento aziendale, consentendo confronti mirati tra le diverse aree dell'organizzazione.

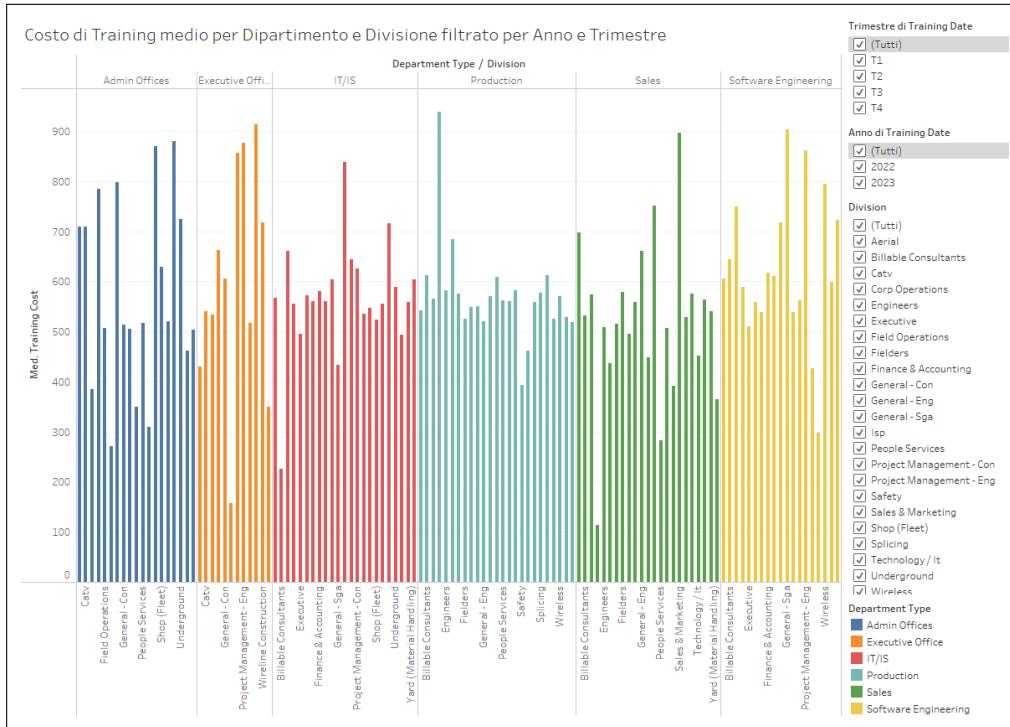


Figura 3.5: Analisi sul costo della formazione

3.4.1 Analisi

In questo caso specifico, Figura 3.6, l'analisi è stata effettuata tenendo conto di tutti i dipartimenti e dei diversi anni di formazione che vanno dal 2022 al 2023, divisi in 4 trimestri, focalizzandosi sulle divisioni aziendali *Engeneers*, *Executive*, *Field Operation*, *General- Eng*, *Saferty*, *Sales & Marketing* e *Tecnology/It*.

Dai dati emerge che il costo medio per la formazione varia significativamente tra le divisioni. Ad esempio, la divisione Executive Office registra i valori più alti, circa 900 €, mentre ci sono punti in cui il costo medio per la formazione è molto basso, circa 400€. Questo grafico consente di identificare le aree con i maggiori investimenti in formazione, evidenziando le priorità aziendali e le possibili opportunità di ottimizzazione.

La presenza di costi rilevanti in divisioni trasversali come *Technology / IT*, *Engeneers* e *Safety* indica che la formazione in queste aree è considerata cruciale per più dipartimenti. Inoltre, l'andamento dei costi potrebbe riflettere necessità aziendali specifiche, come

l'aggiornamento su nuove tecnologie, la sicurezza sul lavoro o il miglioramento delle competenze manageriali e commerciali.

Questa analisi fornisce spunti utili per ottimizzare la pianificazione delle risorse destinate alla formazione, garantendo un equilibrio tra investimenti e necessità aziendali.

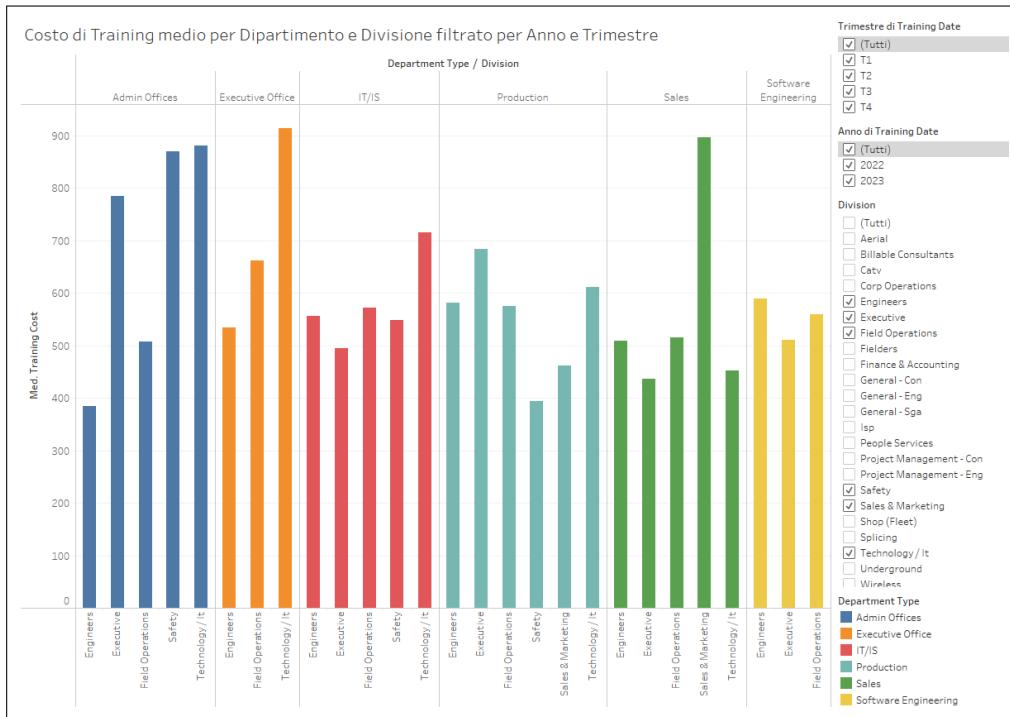


Figura 3.6: Analisi sul costo della formazione filtrato per divisione

Inoltre, abbiamo effettuato un ulteriore analisi, Figura 3.7, filtrando per anno 2022 e i dipartimenti *IT/IS*, *Production* e *Software Engineering* con le relative divisioni aziendali.

Dai dati emergono che il dipartimento *IT/IS* mostra un grande variabilità nei costi di formazione con picchi di circa 900€ e in alcune divisioni i costi sono all'incirca pari a 100€. Questo potrebbe indicare una maggiore necessità formazione specialistica. Il dipartimento *Software Engineering* evidenzia costi più elevati con picchi superiori 900€. Questo potrebbe riflettere l'importanza di formazione avanzata per competenze altamente tecniche. Da ciò l'analisi suggerisce la necessità di valutare l'efficacia degli investimenti nella formazione e di esplorare possibili ottimizzazioni, garantendo al contempo che le competenze critiche per ogni dipartimento siano adeguatamente sviluppate.

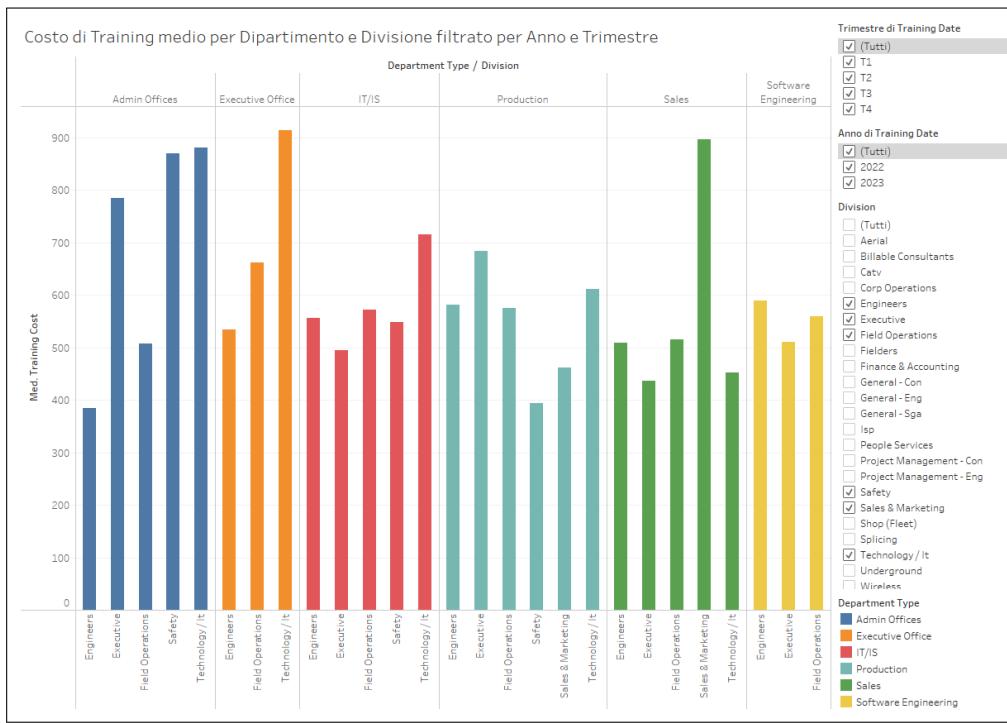


Figura 3.7: Analisi sul costo della formazione filtrato per anno e dipartimento

3.5 Dashboard 3: Analisi del successo della formazione per fascia retributiva

Il terzo foglio della nostra analisi in Tableau fornisce una visone dettagliata della tipologia della formazione per fascia retributiva dei dipendenti, con particolare attenzione alla percentuale di successo per ogni tipo di formazione.

Per analizzare la distribuzione dei dipendenti in base alla loro formazione, sono state utilizzate diverse rappresentazioni grafiche:

- Grafico a barre: mostra la distribuzione dei dipendenti in base alla fascia retributiva (Pay Zone) e alla divisione di cui fa parte, differenziando con diversi colori le tipologie di formazione. Inoltre, è stato utilizzato per rappresentare la percentuale di successo nei vari dipartimenti a seconda della tipologia del corso di formazione.
- Grafico a torta: rappresenta la ripartizione dei dipendenti, classificati secondo la loro formazione, in base alla tipologia di contratto.

Nella dashboard sono anche presenti diversi filtri, tra cui:

- *Division*: consente di selezionare specifiche divisioni aziendali all'interno dei dipartimenti.

- *Department Type*: permette di analizzare i dati filtrando per specifico dipartimento aziendale, consentendo confronti mirati tra le diverse aree dell’organizzazione.
- *Training Type*: consente la possibilità di selezionare la tipologia di formazione

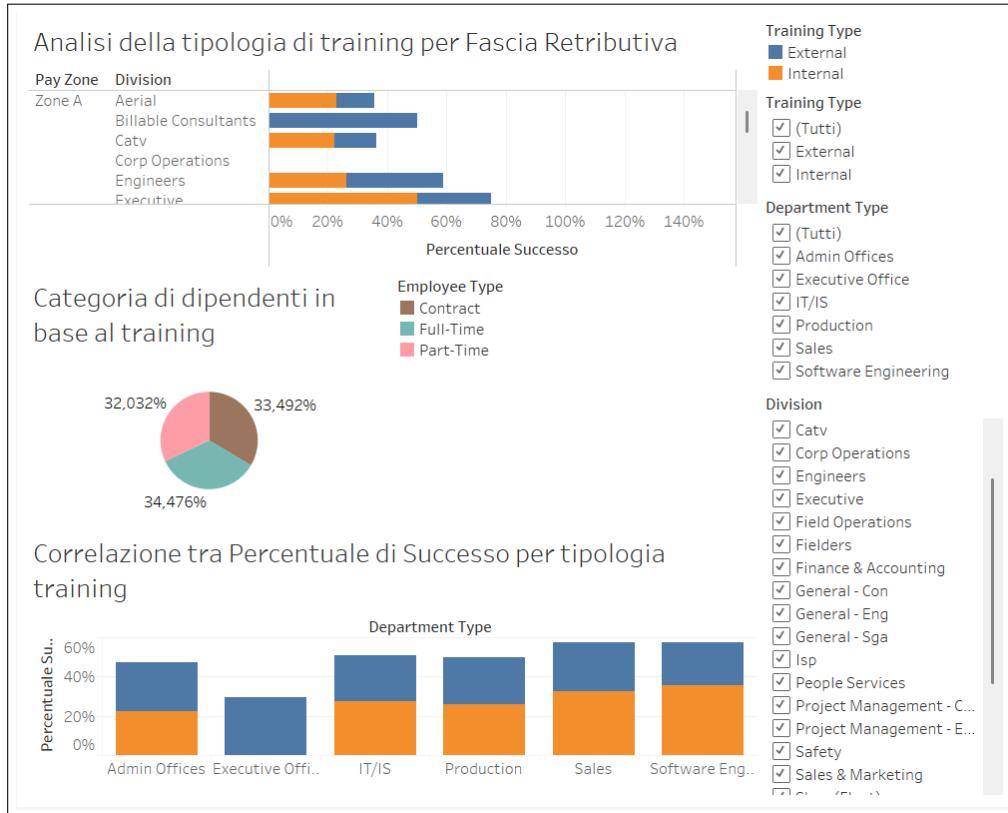


Figura 3.8: Analisi del successo della formazione per fascia retributiva

3.5.1 Analisi

In questo caso specifico, Figura 3.8, l’analisi è stata effettuata selezionando la formazione come Interna (Internal) e focalizzandosi sui dipartimenti *IT/IS*, *Admin Offices* e *Executive Office* con le relative divisioni e le aree aziendali.

Dai dati emerge che la percentuale di successo di un dipendente di riuscire ad arrivare alla fascia retributiva più alta (Zone A) non è superiore al 50%. In alcune divisioni, come General- Spa o People Services, notiamo che risulta impossibile arrivare alla fascia con il salario più alto, mentre in altre sfiora il 20%. Se andiamo ad analizzare la percentuale di successo per dipartimento riscontriamo delle percentuali ancora più basse fino ad un massimo del 30% in *IT/IS* e *Admin Offices*, mentre in *Executive Office* la percentuale di successo è nulla perché la formazione è soltanto esterna.

Inoltre, nel grafico a torta che mostra la distribuzione dei dipendenti notiamo che soltanto il 28.6% ha un contratto Full-Time e il 36.2% ha un contratto Part-Time, il restante 35.2% rappresenta i dipendenti che hanno un contratto a tempo determinato. Questa analisi mette in evidenza che la formazione interna risulta più diffusa tra i dipendenti che hanno un contratto temporaneo, che rappresentano una percentuale maggiore rispetto a quelli Part-Time e Full-Time. Questo suggerisce che l'azienda investa principalmente nella crescita professionale dei dipendenti che si sono appena inseriti, per garantire continuità e per migliorare le loro competenze o prepararli a ruoli più stabili all'interno dell'azienda.

Un altro aspetto rilevante è che la percentuale di successo della formazione interna rispetto a quella esterna non presenta evidenti differenze, l'azienda potrebbe valutare una strategia di ottimizzazione, aumentando l'offerta di corsi interni per ridurre i costi e mantenere la qualità dell'apprendimento. Nonostante ciò, alcuni dipartimenti riescono comunque a ottenere buoni risultati con la sola formazione interna. In particolare, i dipartimenti *Admin Offices* e *IT/IS* mostrano una discreta percentuale di successo, segnalando che, in questi ambiti, il training interno è probabilmente ben strutturato ed efficace nel garantire un livello di apprendimento adeguato.

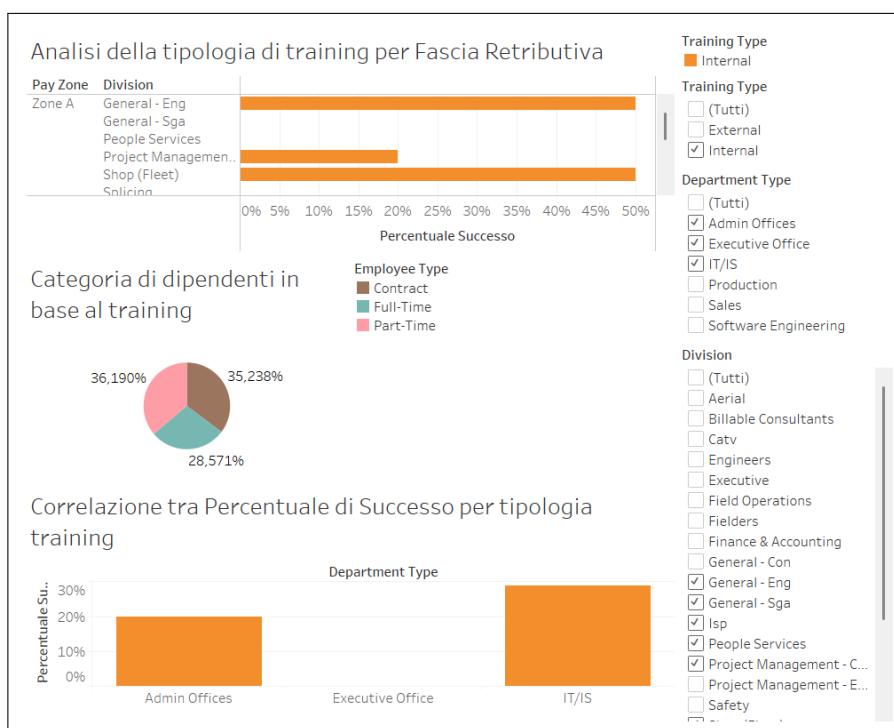


Figura 3.9: Analisi del successo della formazione per fascia retributiva filtrato

3.6 Dashboard 4: Analisi sul risultato della formazione

Il quarto foglio della nostra analisi in Tableau, Figura 3.10, fornisce un'analisi sui costi della formazione di ogni dipendente in base all'esito.

La visualizzazione è organizzata attraverso un filtro gerarchico che consente di selezionare i dati in base a *DepartmentType* e *Division*, permettendo un'analisi mirata su specifici gruppi di dipendenti.

In alto troviamo un grafico a dispersione che mostra la relazione la media dei costi di formazione con l'identificativo di ogni dipendente. Nella parte centrale è presente un diagramma a barre che mostra la relazione l'esito della formazione in base al dipartimento. Gli esiti della formazione sono quattro:

- *Completed*: Il dipendente ha completato il corso di formazione, ma senza necessariamente superarlo con successo.
- *Passed*: Il dipendente ha completato con successo la formazione, raggiungendo i criteri di valutazione richiesti.
- *Incomplete*: Il dipendente ha iniziato la formazione ma non l'ha terminata.
- *Failed*: Il dipendente ha completato la formazione ma non ha raggiunto il punteggio minimo richiesto per superarla.

Questo diagramma consente di individuare come l'esito della formazione dei dipendenti si e come si distribuisce nei vari dipartimenti e divisioni aziendali.

Infine, nella parte inferiore possiamo osservare un grafico lineare che descrive l'andamento temporale degli esiti di formazione in relazione al costo medio, aiutando la gestione dei budget per ottimizzare la spesa formativa.

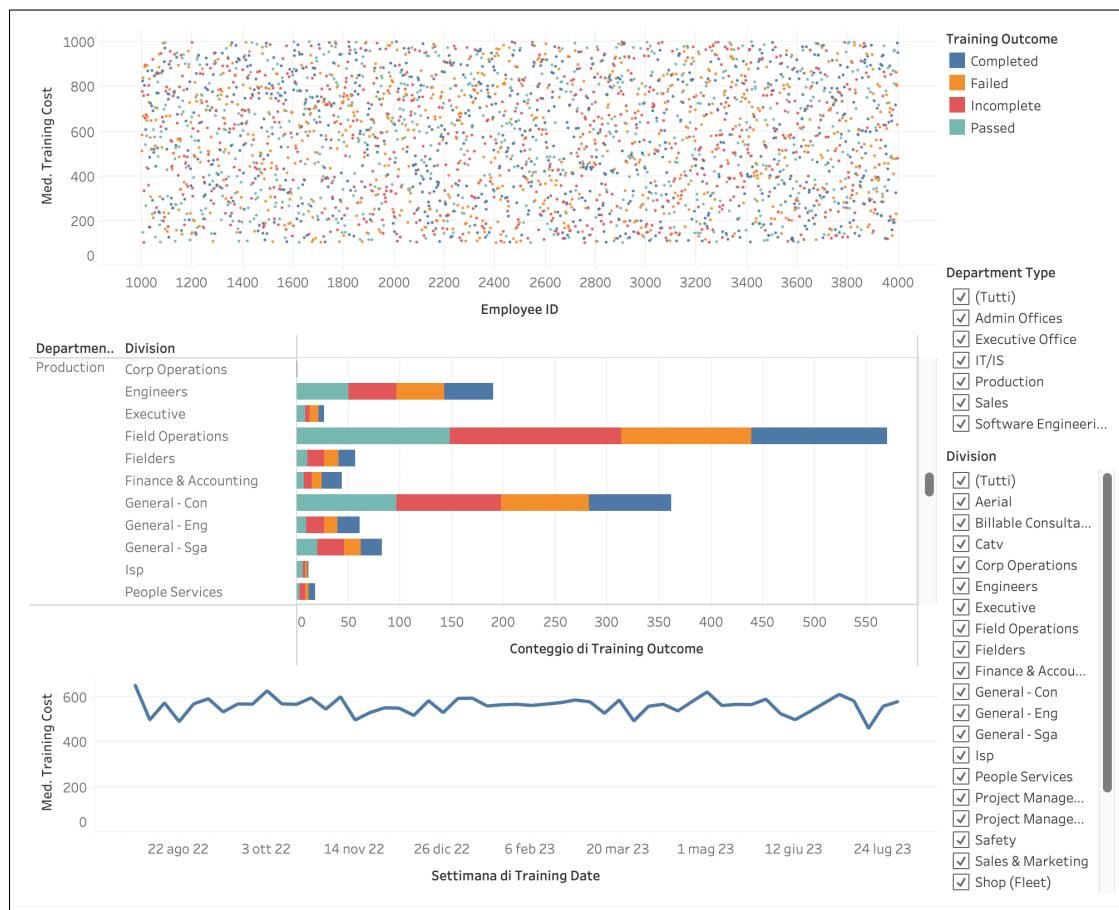
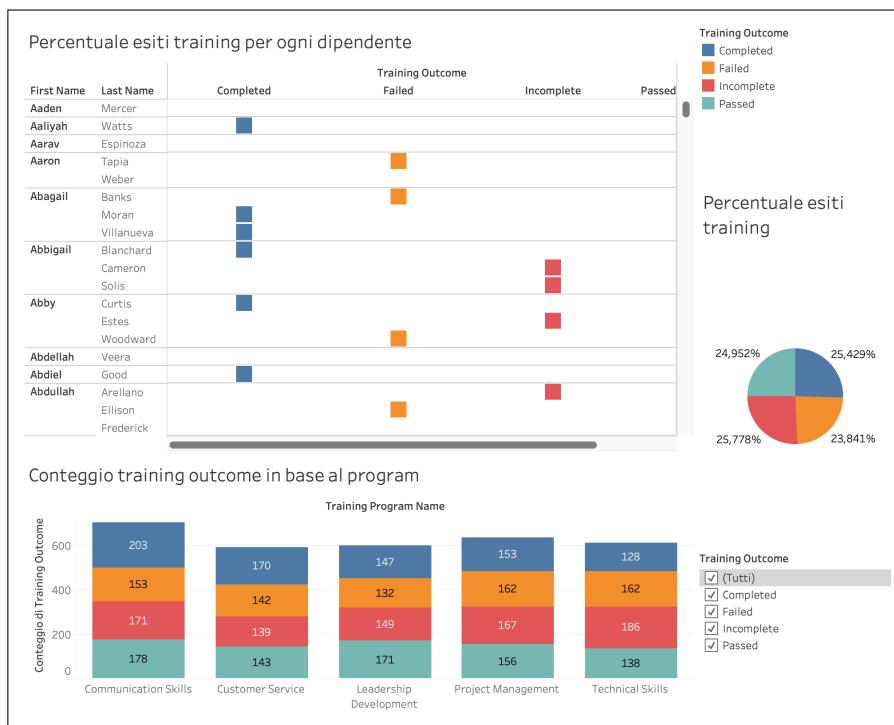


Figura 3.10: Analisi sul costo della formazione

In quest'ultimo foglio della nostra analisi in Tableau, Figura 3.11, forniamo un'analisi sulle percentuali degli esiti della formazione di ogni dipendente. La visualizzazione è organizzata attraverso un filtro che consente di selezionare i dati in base all'esito della formazione.

In alto troviamo una heatmap (mappa di calore) a matrice che mette in relazione ogni singolo dipendente, identificato con nome e cognome, con il rispettivo esito della formazione. Questo grafico permette di visualizzare rapidamente la distribuzione degli esiti per ogni dipendente, utile per un confronto immediato tra diversi dipendenti. Nella parte di destra abbiamo un grafico a torta che ci evidenza le percentuali di ognuno dei quattro esiti, utile per un confronto immediato tra i vari esiti. Le percentuali mostrano che nessun esito domina in modo schiacciante, suggerendo che ci sia una distribuzione bilanciata tra successi (Completed e Passed) e insuccessi (Incomplete e Failed).

Infine, in basso troviamo un grafico a barre che mostra il conteggio degli esiti in base alla tipologia del corso di formazione. Questo grafico ci permette di analizzare sia il totale dei partecipanti per corso, sia la distribuzione degli esiti.

**Figura 3.11:** Analisi sul risultato della formazione

3.6.1 Analisi

In questo caso specifico, Figura 3.12, l'analisi è stata effettuata selezionando come tipologia di esito della formazione *Completed* e *Passed* in modo da osservare il confronto dei prezzi dei corsi di formazione tra i vari dipartimenti e le differenti divisioni.

Dai dati emerge che osservando il grafico a dispersione posto in alto non si hanno variazioni sulla media del costo di formazione. Nel grafico a barre si osserva che il numero di esiti registrati è significativamente inferiore rispetto all'analisi principale. Questo indica che solo una piccola parte dei dipendenti ottiene un esito positivo rispetto al totale dei partecipanti. Tale risultato evidenzia la necessità di migliorare i corsi di formazione, rendendoli più efficaci e aumentando il tasso di successo. Infine, nell'ultimo grafico osserviamo che non ci sono sostanziali differenze con quello non filtrato, quindi sta a significare che il costo medio è simile tra i due grafici. Questo indica che si sta spendendo molto su corsi che non portano ai risultati sperati per via di un numero di successi molto basso e un alto valore del costo medio di formazione.



Figura 3.12: Analisi costi della formazione

Anche in questo caso, Figura 3.13, l'analisi è stata effettuata selezionando come tipologia di esito della formazione *Completed* e *Passed* in modo da osservare il confronto delle percentuali di successo dei corsi di formazione tra i vari dipartimenti e le differenti divisioni.

Dai dati emerge che il conteggio degli esiti di corsi di formazione è inferiore. Infatti nel grafico a colonne il valore del conteggio passa da circa 600 a circa 400. Questo indica che si ha un maggior numero di esiti positivi, permettendo di identificare quali dei corsi sono stati più efficaci, identificando quali programmi di formazione funzionano meglio e quali richiedono miglioramenti. Per entrare più in dettaglio nel grafico in alto troviamo i dipendenti, identificati con nome e cognome e i rispettivi risultati. Inoltre, mettendo a confronto le due tipologie di risultati emerge che la percentuale di esiti *Completed*, il 50,5%, è leggermente superiore rispetto a quella dei *Passed*, il 49,5%. Da questo capiamo che tutti riescono a completare il corso di formazione, ma non tutti ottengono un vero e proprio superamento.

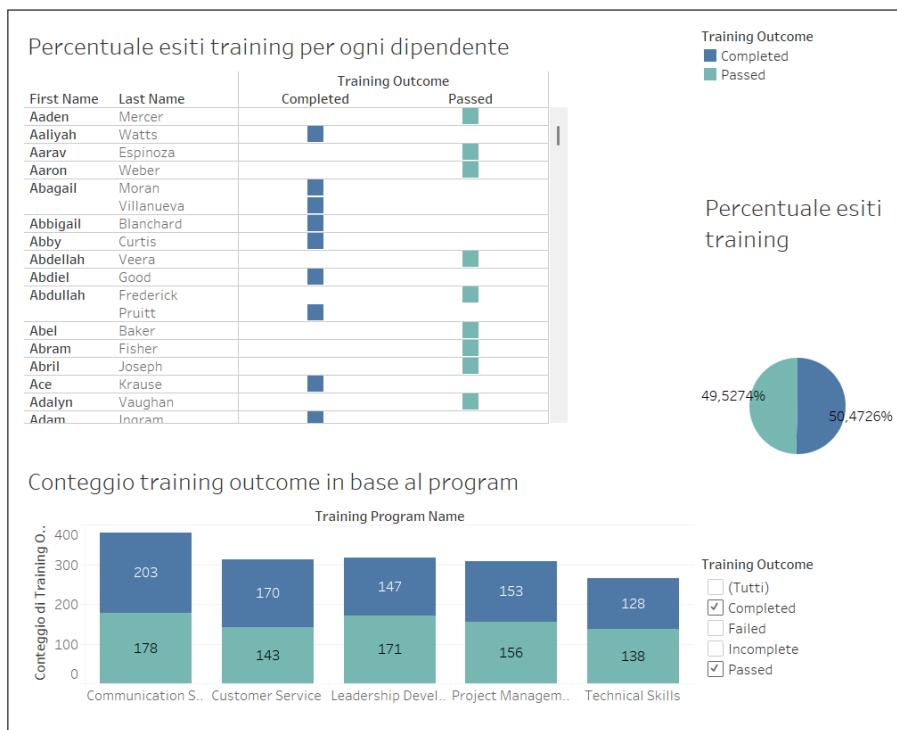
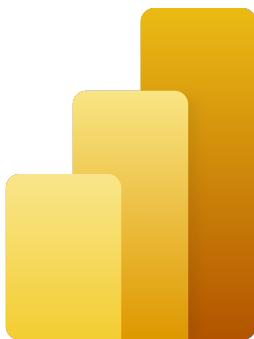


Figura 3.13: Analisi percentuali di successo della formazione

CAPITOLO 4

Power BI



4.1 Introduzione

Power BI è una piattaforma di Business Intelligence sviluppata da Microsoft che consente alle aziende di connettere, analizzare e visualizzare i dati provenienti da diverse fonti, trasformandoli in insight utili per il processo decisionale. Grazie a strumenti avanzati di analisi, Power BI permette agli utenti di creare dashboard interattive e report dinamici.

La piattaforma si compone di diversi strumenti principali:

- *Power BI Desktop*: un'applicazione per la creazione di report e analisi approfondite.
- *Power BI Service*: una piattaforma cloud per la condivisione e collaborazione sui report.
- *Power BI Mobile*: un'app per visualizzare i report su dispositivi mobili.
- *Power BI Gateway*: un servizio per connettere fonti dati locali con il cloud.

Uno dei punti di forza di Power BI è la sua profonda integrazione con l'ecosistema Microsoft, in particolare con Excel, Azure e SQL Server, oltre alla possibilità di connettersi a numerose altre fonti di dati, come servizi cloud e database aziendali.

4.2 Come funziona

Il processo di utilizzo di Power BI si articola in tre fasi principali:

1. Caricamento dei dati: Power BI consente di connettersi a un'ampia varietà di fonti dati, tra cui file Excel, database SQL, servizi cloud come Google Analytics e Salesforce, oltre a numerosi API e fonti online.
2. Modellazione e trasformazione: Attraverso l'editor Power Query, gli utenti possono pulire e trasformare i dati, creare relazioni tra tabelle e definire misure calcolate con il linguaggio DAX (Data Analysis Expressions).
3. Creazione di dashboard e report: Power BI offre una vasta gamma di visualizzazioni per rappresentare i dati in modo chiaro e interattivo.

Alcuni degli oggetti principali disponibili in Power BI includono:

- *Grafici a barre e colonne*: utili per confrontare valori tra diverse categorie.
- *Grafici a torta e ad anello*: mostrano la distribuzione percentuale dei dati.
- *Tabelle e matrici*: consentono di visualizzare dati tabellari con funzioni avanzate di ordinamento e filtri.
- *Carte KPI*: mostrano metriche chiave con indicatori visivi di performance.
- *Mappe geospaziali*: rappresentano dati geografici utilizzando Bing Maps.
- *Slicer e filtri*: permettono di interagire con i dati e affinare l'analisi.

Dopo aver creato le visualizzazioni, queste vengono organizzate all'interno di *pagine* (o fogli) all'interno di un report Power BI. Ogni pagina può contenere diverse visualizzazioni interattive, permettendo un'esplorazione approfondita dei dati.

Nella nostra analisi, abbiamo realizzato i seguenti fogli:

- Foglio 1: Analisi delle assunzioni.
- Foglio 2: Analisi delle dimissioni.
- Foglio 3: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva.
- Foglio 4: Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione.

4.3 Foglio 1: Analisi delle assunzioni

Il primo foglio dell' analisi in Power BI fornisce una visione dettagliata sull'andamento delle assunzioni nel tempo, con particolare attenzione alla variazione annuale e alla possibilità di effettuare previsioni per il futuro. Il report include diverse visualizzazioni interattive che permettono di esplorare i dati a più livelli di dettaglio e confrontare le assunzioni rispetto agli anni precedenti.

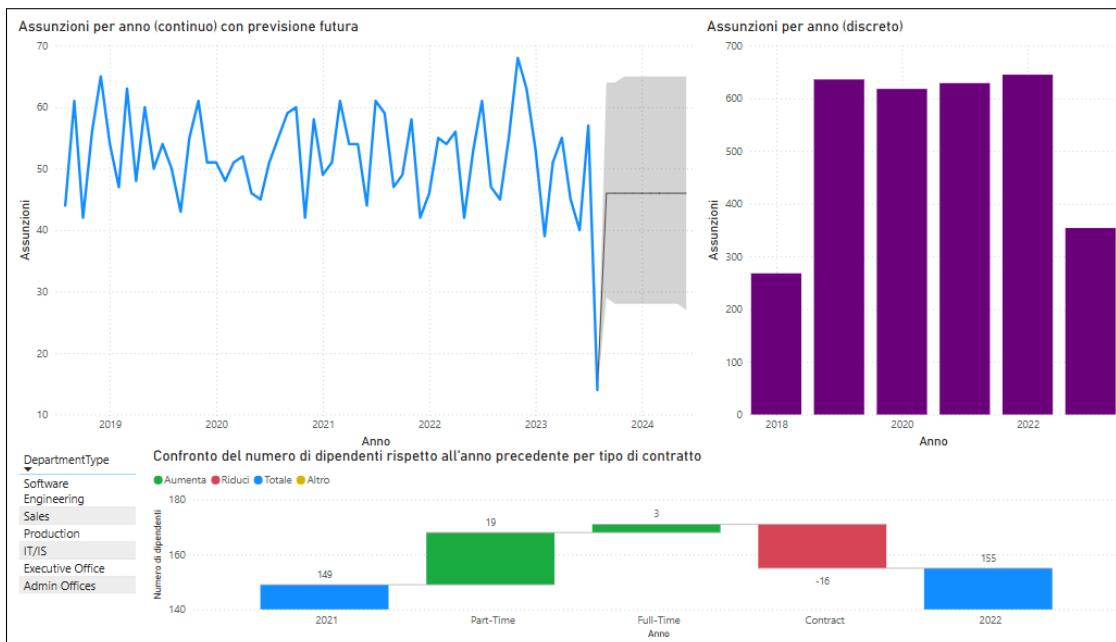


Figura 4.1: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva.

Per analizzare l'evoluzione delle assunzioni, sono state utilizzate due diverse rappresentazioni grafiche:

- Grafico a linee: mostra l'andamento continuo delle assunzioni nel tempo, permettendo di approfondire il dato a livelli più dettagliati (trimestre, mese, ecc.). Inoltre, include una previsione delle assunzioni future basata sulle tendenze storiche.
- Grafico a colonne: fornisce una rappresentazione discreta delle assunzioni per anno, offrendo un confronto chiaro tra i vari periodi.

Un'ulteriore analisi presente nel foglio riguarda il confronto tra il numero totale di dipendenti assunti e quello registrato nell'anno precedente. L'anno di riferimento è impostato al 2022, e il confronto è suddiviso per tipo di contratto, distinguendo tra: *Full-time*, *Part-time* e *Contract*.

Per garantire un'analisi più mirata, il report offre la possibilità di filtrare tutte le visualizzazioni in base al tipo di dipartimento, consentendo di analizzare specifiche aree aziendali e identificare eventuali trend differenziati per settore.

4.3.1 Analisi

Dall'analisi emerge che il numero di assunzioni ha mostrato oscillazioni significative nel corso degli anni, senza seguire un andamento lineare o prevedibile. Un aspetto interessante è che l'andamento delle assunzioni non è perfettamente allineato con quello delle dimissioni (4.7), il che potrebbe suggerire dinamiche aziendali indipendenti dai flussi di turnover del personale. Infatti ci si aspetterebbe che tra il 2023 e il 2024 ci siano numerose assunzioni dovute alle numerose dimissioni che verranno mostrate nelle sezioni successive. Questo potrebbe essere dovuto ad un andamento negativo dell'azienda. La previsione effettuata con un intervallo di confidenza al 95% risulta essere piatta, in linea con la media, in quanto non si è riuscito a trovare una relazione tra le assunzioni e il tempo. Per effettuare la prima analisi è stato necessario impostare come continuo il numero di assunzioni fatte. A destra si mostra il numero di assunzioni per anno discretizzato con un grafico colonne.

Un ulteriore elemento di analisi riguarda la distribuzione dei dipendenti per anno. Per ottenere un confronto più dettagliato, è stata inserita la misura:

```
Assunzioni_Per_Anno = COUNT(Sheet1[StartDate])
```

che, inserita sulle ascisse, consente di esaminare il numero complessivo di dipendenti per ciascun anno di riferimento.



Figura 4.2: Analisi delle assunzioni nel tempo, nel dipartimento Produzione per tipo di contratto.

Inoltre, il report permette di filtrare i dati per dipartimento, consentendo di focalizzare l'attenzione su specifiche aree aziendali. Ad esempio, nella Figura 4.2 è stato applicato

un filtro sul dipartimento di *Produzione*, mettendo in evidenza i seguenti cambiamenti rispetto all'anno precedente:

- Aumento dei contratti Part-time di +15.
- Diminuzione dei contratti Contract di -9.
- Aumento dei contratti Full-time di +3.

Queste variazioni si traducono in un saldo positivo complessivo di +9 dipendenti nel dipartimento di produzione.

Per andare ancora più in dettaglio si potrebbero visualizzare ad esempio solo le dimissioni dei dipendenti a Contratto nell'arco temporale nel dipartimento produzione come visualizzato in Figura 4.3, andando a vedere quante sono state le assunzioni e perché il numero risulta negativo.

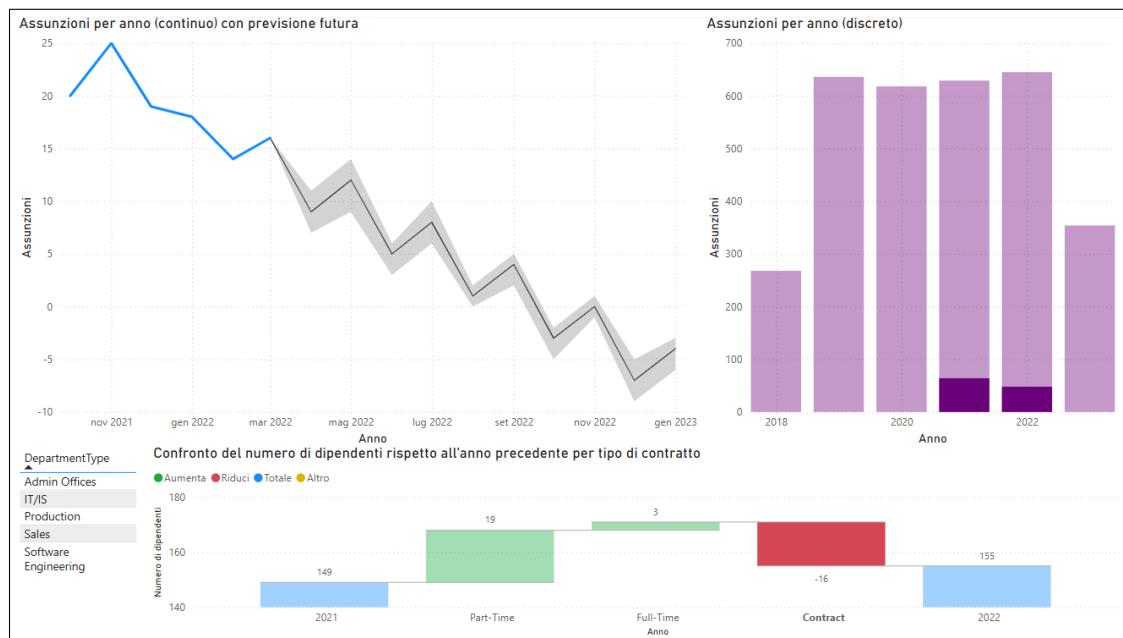


Figura 4.3: Analisi delle assunzioni nel tempo, nel dipartimento Produzione per dipendenti a Contratto

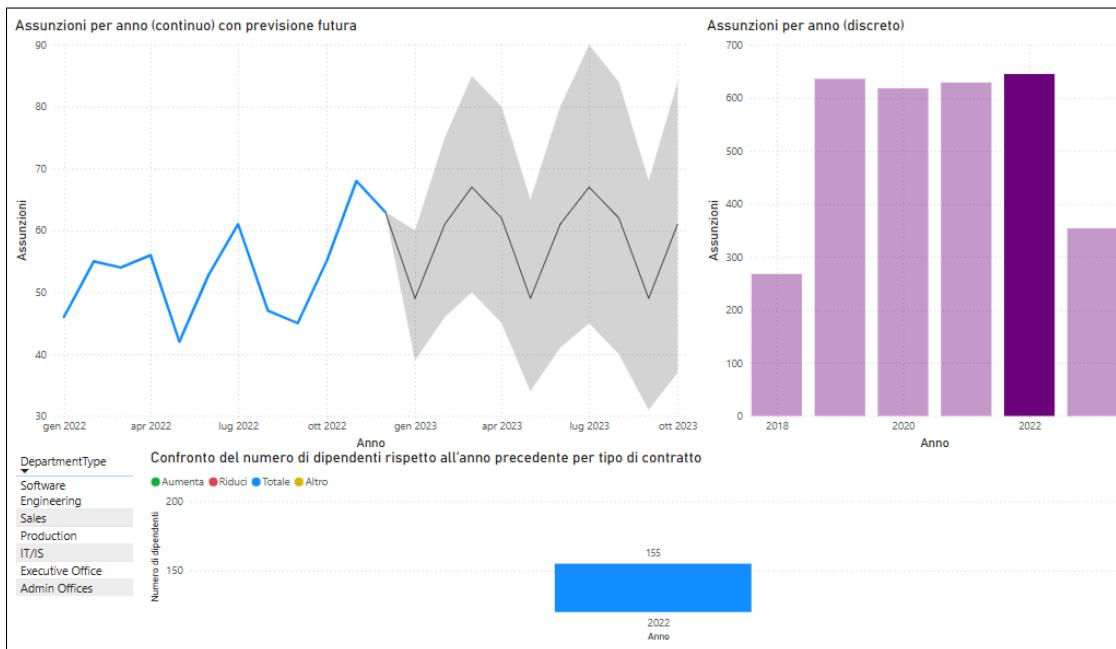


Figura 4.4: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva nell'anno 2022.

Il report permette anche di selezionare un singolo anno su cui effettuare l'analisi. Ad esempio, nella Figura 4.4 si seleziona l'anno 2022 ed è possibile notare come emerge un pattern nelle assunzioni che evidenzia un andamento più o meno regolare. Questo permette di fare delle analisi a livello annuale con più precisione, anche se poi a livello globale tali previsioni non vengono rispettate. Naturalmente l'ultimo grafico che rappresentante il numero di dipendenti assunti rispetto all'anno precedente mostra unicamente i dipendenti assunti in quell'anno.

4.4 Foglio 2: Analisi delle dimissioni

Il secondo foglio della nostra analisi in Power BI è dedicato allo studio delle dimissioni dei dipendenti all'interno dell'azienda 4.5. Questo report permette di esaminare la percentuale di dimissioni per divisione, confrontare il numero di dipendenti attualmente in azienda con quelli che la hanno lasciata e analizzare l'andamento storico delle dimissioni con una previsione basata sui dati passati.

L'analisi delle dimissioni risulta utile per identificare eventuali problematiche legate alla gestione del personale. Grazie all'integrazione dei diversi grafici e strumenti di filtro, questo foglio consente di:

- Individuare le divisioni con il tasso di dimissioni più elevato, facilitando l'identificazione di possibili criticità.
- Confrontare il numero di dipendenti attivi e dimessi, valutando l'impatto del turnover¹ sull'azienda.
- Visualizzare la durata media di un dipendente all'interno dell'azienda.
- Analizzare l'andamento delle dimissioni nel tempo e prevederne l'evoluzione futura, fornendo un supporto concreto per la pianificazione strategica.

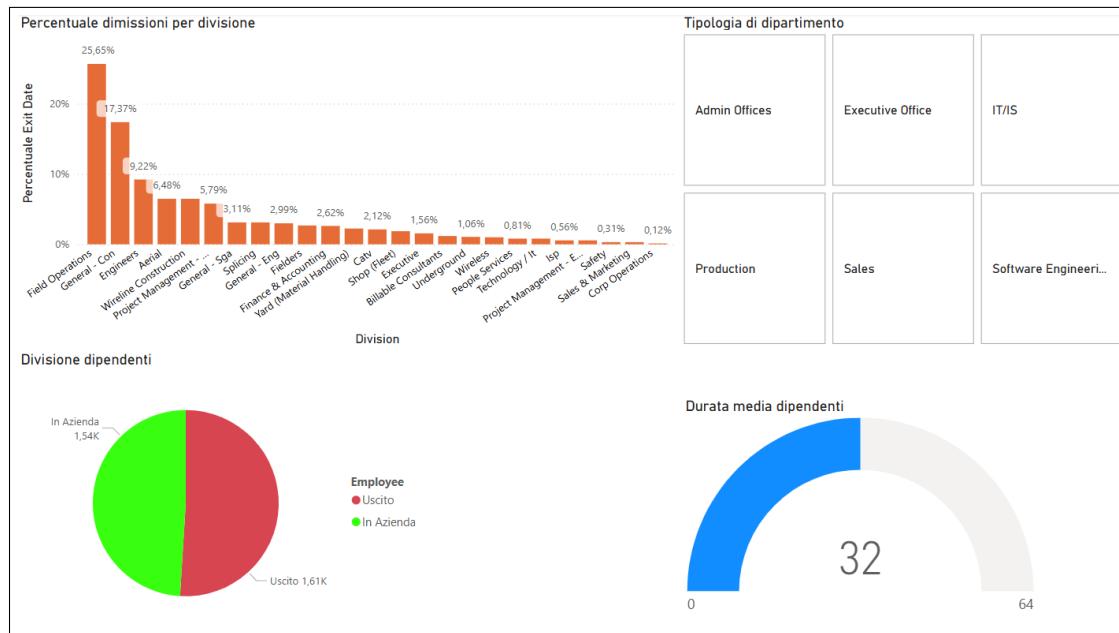


Figura 4.5: Analisi delle dimissioni.

¹Il turnover in ambito aziendale si riferisce al tasso con cui i dipendenti lasciano un'azienda e vengono sostituiti da nuovi assunti.

4.4.1 Analisi

La prima visualizzazione del report è rappresentata da un grafico a colonne che mostra la percentuale di dimissioni per ciascuna divisione aziendale. Questa analisi consente di individuare le aree con il tasso di turnover più elevato, fornendo un primo spunto per approfondire le cause delle dimissioni in specifiche divisioni.

Inoltre, come possibile vedere in Figura 4.6, è possibile affinare l'analisi selezionando uno specifico dipartimento.

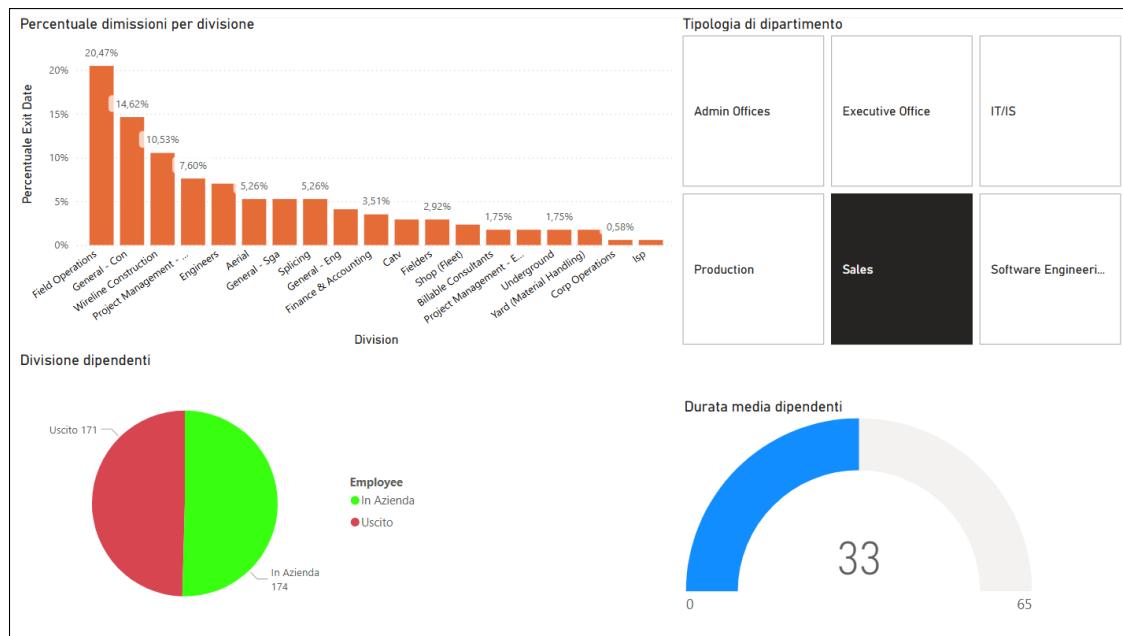


Figura 4.6: Analisi delle dimissioni per dipartimento Sales.

In questo modo, l'utente può concentrarsi su una determinata area aziendale e confrontare i tassi di dimissioni tra le diverse divisioni interne. Si può notare infatti come ci sia un problema significativo all'interno della divisione 'Field Operations', che rappresenta la colonna con il maggior numero di dimissioni. Ci si aspetta che sia anche la divisione con il maggior numero di problematiche e che dunque vada approfondita la questione.

Per comprendere meglio l'impatto delle dimissioni sull'organico aziendale, il secondo grafico del report è un grafico a torta che rappresenta la proporzione tra i dipendenti attualmente in azienda e quelli dimessi. Questa visualizzazione offre una panoramica immediata della distribuzione del personale e consente di valutare il peso delle dimissioni rispetto al totale della forza lavoro.

L'ultima sezione del foglio (in Figura 4.7) è dedicata all'analisi dell'andamento nel

tempo delle dimissioni. Attraverso un grafico a linee, è possibile osservare l’evoluzione del numero di dipendenti dimessi nel corso degli anni, mettendo in evidenza eventuali trend o cambiamenti significativi nelle politiche aziendali o nelle condizioni di lavoro. Per fornire una visione più completa è stata inoltre implementata una previsione delle dimissioni future utilizzando le funzionalità avanzate di Power BI. La previsione è stata elaborata con un livello di confidenza del 95%, permettendo di stimare l’andamento futuro delle dimissioni e di individuare possibili scenari di evoluzione dell’organico aziendale. Come si può notare la tendenza che segue il numero di dimissioni è in crescita, ma nell’ultimo anno si è verificato un picco negativo. Come evidenziano le previsioni però, ci si aspetterebbe che sia stata una casualità dato che il trend degli anni precedenti risultava crescente. Al contrario potrebbe anche essere dovuto ad un cambio gestione o un cambiamento radicale nell’azienda che ha portato ad arrestare il fenomeno di crescita. Per questo l’intervallo di confidenza al 95%, essendo molto ampio, riesce a rappresentare entrambe le possibilità. Questa analisi però risulta essere in contraddizione con quanto riportato nel KPI rappresentante la durata media di un dipendente nell’azienda. I valori che ci indica il KPI fanno riferimento alla data odierna, mentre i dati relativi a dimissioni o licenziamenti si fermano nel 2023. Questo ha reso l’analisi irrilevante, ma comunque utile a scopo didattico. Per farla è stato necessario aggiungere la misura:

```
DurataMediaMesi =
AVERAGEX(
Sheet1,
DATEDIFF(
Sheet1[StartDate],
IF(
ISBLANK(Sheet1[ExitDate]),
TODAY(),
Sheet1[ExitDate]
),
MONTH
)
)
```

L’ultimo punto calcola la percentuale di Turnover attraverso la misura creata:

```
Turnover = CALCULATE(
COUNT(Sheet1[Employee ID]),
NOT(ISBLANK(Sheet1[ExitDate])))
)
```

In pratica la formula DAX restituisce il numero totale di dipendenti che hanno lasciato l’azienda, basandosi sul fatto che la loro colonna ExitDate non sia vuota. Successivamente, dividendo per il totale dei dipendenti:

```
Percentuale_Turnover=  
DIVIDE([Turnover], [Totale_Dipendenti], 0)
```

è stata creata la misura rappresentante le percentuale del Turnover. Premendo poi in una data specifica nella sequenza temporale è possibile vedere quale sia la percentuale di turnover in quella specifica data.

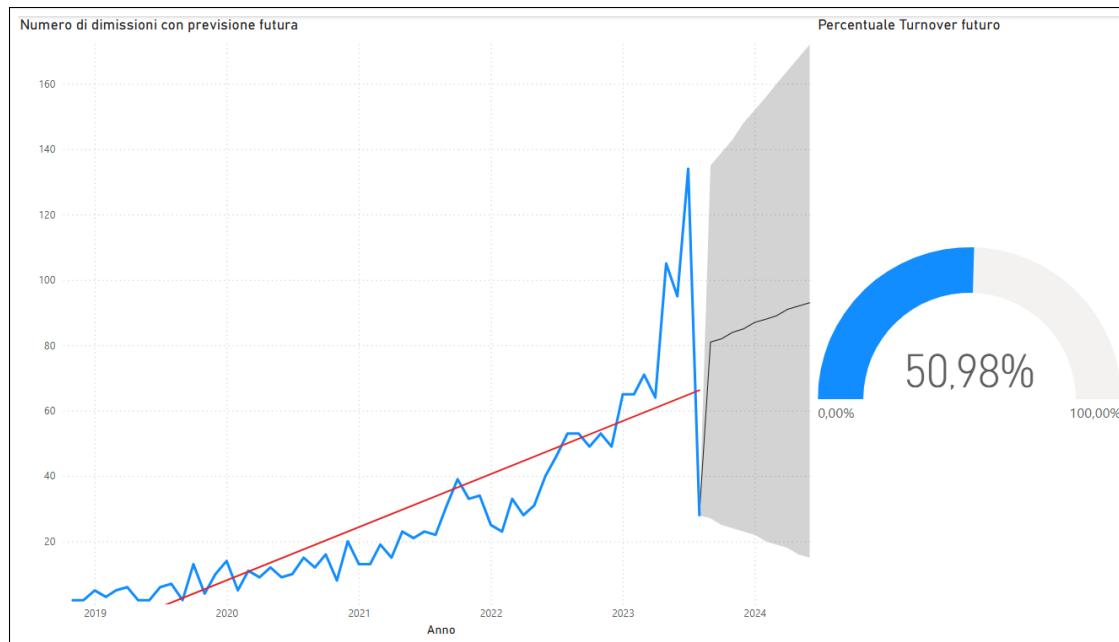


Figura 4.7: Analisi delle dimissioni per durante gli anni con previsione futura

4.5 Foglio 3: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva

Il terzo foglio della nostra analisi in Power BI offre una panoramica dettagliata sulla distribuzione dei dipendenti all'interno dell'azienda, suddivisi per tipo di impiego e fascia retributiva. Il report è altamente interattivo e consente di esplorare i dati applicando filtri dinamici basati su genere ed etnia oltre che alla localizzazione geografica.

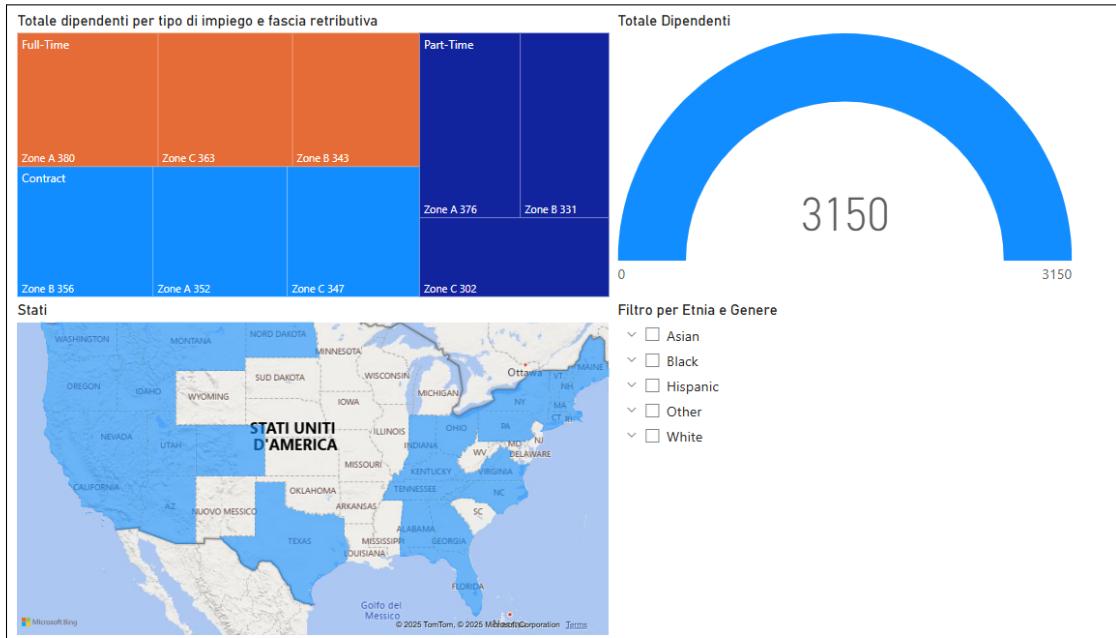


Figura 4.8: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva.

Nella prima Figura (4.8) è possibile notare il quadro generale e completo della situazione. Per avere il numero totale dei dipendenti si è aggiunta la misura:

```
Totale_Dipendenti = COUNT(Sheet1[Employee ID])
```

che è risultata molto utile anche nelle analisi successive. Il tipo di impiego è suddiviso in tre categorie: *Full-time*, *Part-time* e *Contract*.

Il tipo di fascia retributiva comprende: *ZoneA* (la più redditizia), *ZoneB* e *ZoneC* (la meno redditizia).

Il genere è classificato tra: *Male* e *Female*.

L'etnia comprende le seguenti categorie: *White* *Hispanic* *Black* *Asian* *Other*.

Grazie all'integrazione dei diversi grafici e strumenti di filtro, questo foglio consente di:

- Analizzare la distribuzione dei dipendenti identificando eventuali tendenze o differenze tra i gruppi di genere, etnia, contratti o salari diversi.

4.5 – Foglio 3: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva 47

- Confrontare la distribuzione della forza lavoro per stabilire eventuali strategie di espansione dell'azienda.

4.5.1 Analisi

Dai dati analizzati emerge subito un sostanziale equilibrio nella distribuzione dei dipendenti rispetto a diverse forme di categorizzazione. In particolare, non si evidenziano differenze significative tra le fasce retributive (Zona A, Zona B e Zona C), il genere (maschile e femminile) o l'etnia dei dipendenti (White, Hispanic, Black, Asian, Other).

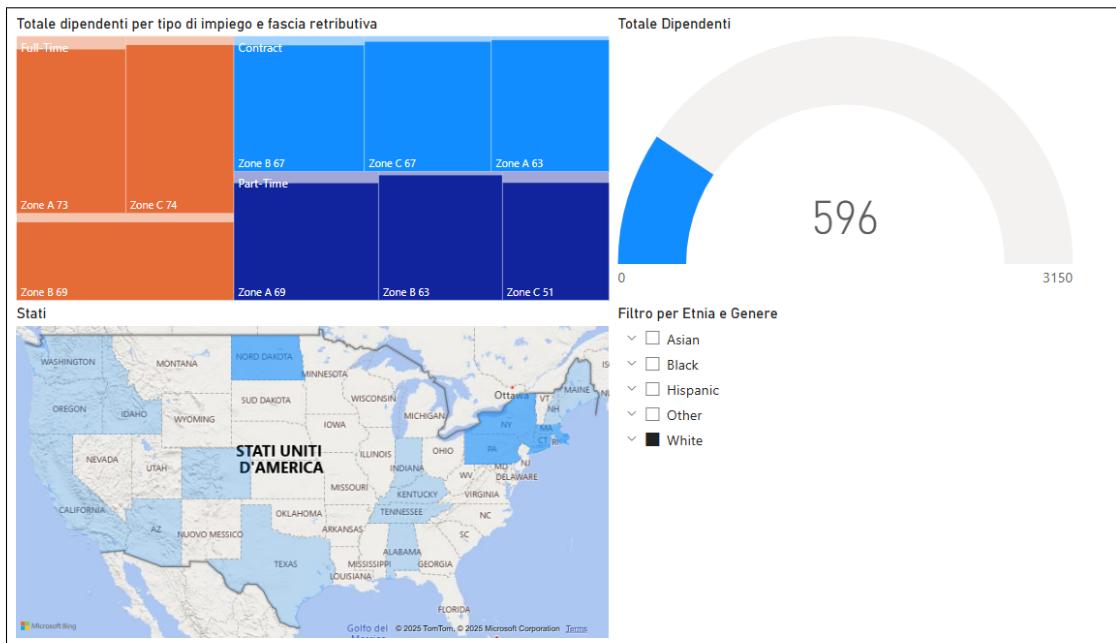


Figura 4.9: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva con filtro per etnia.

Come possibile vedere anche in Figura 4.9, questo equilibrio suggerisce che la politica aziendale relativa all'assegnazione dei salari e alla distribuzione delle opportunità lavorative sia equa e uniforme e si attesti all'incirca sulle 300 persone per zona salariale e tipo di contratto. Non si riscontrano concentrazioni anomale o disparità evidenti tra le categorie analizzate, il che indica che il processo di assunzione e retribuzione è distribuito in modo omogeneo tra i vari gruppi di dipendenti. Al contempo, se da una parte è possibile vedere questa distribuzione come qualcosa di positivo, in realtà suggerisce che i dati siano stati creati artificialmente. Ci si aspetterebbe infatti che la distribuzione della fascia retributiva A sia maggiore tra i dipendenti con il tipo di impiego Full-time rispetto ad un semplice dipendente a contratto. Invece, anche la fascia retributiva risulta essere distribuita uniformemente mostrando delle incongruenze nei dati.

4.5 – Foglio 3: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva48

L'unica variabile che mostra una differenza sostanziale è la localizzazione geografica dei dipendenti. La mappa interattiva utilizzata nel report evidenzia in blu gli stati in cui i dipendenti dell'azienda sono attualmente impiegati. Nella figura 4.10 sono stati selezionati gli stati Americani del nord-est (New York, Connecticut, Rhode Island, Massachusetts, Vermont, Maine e New Hampshire). Dall'analisi della mappa emerge che, sebbene l'azienda abbia una distribuzione su più stati, la maggior parte dei dipendenti si concentra nello stato del Massachusetts. Questo suggerisce che la sede principale dell'azienda si trovi proprio in questo stato o che, comunque, vi sia un polo lavorativo significativamente più grande rispetto alle altre sedi sparse sul territorio. Questa caratteristica potrebbe essere un elemento chiave per future analisi sulla distribuzione della forza lavoro e sulle strategie di espansione dell'azienda, in particolare per comprendere meglio eventuali necessità di ridistribuzione del personale o per analizzare l'impatto della localizzazione geografica sulla struttura occupazionale dell'azienda.

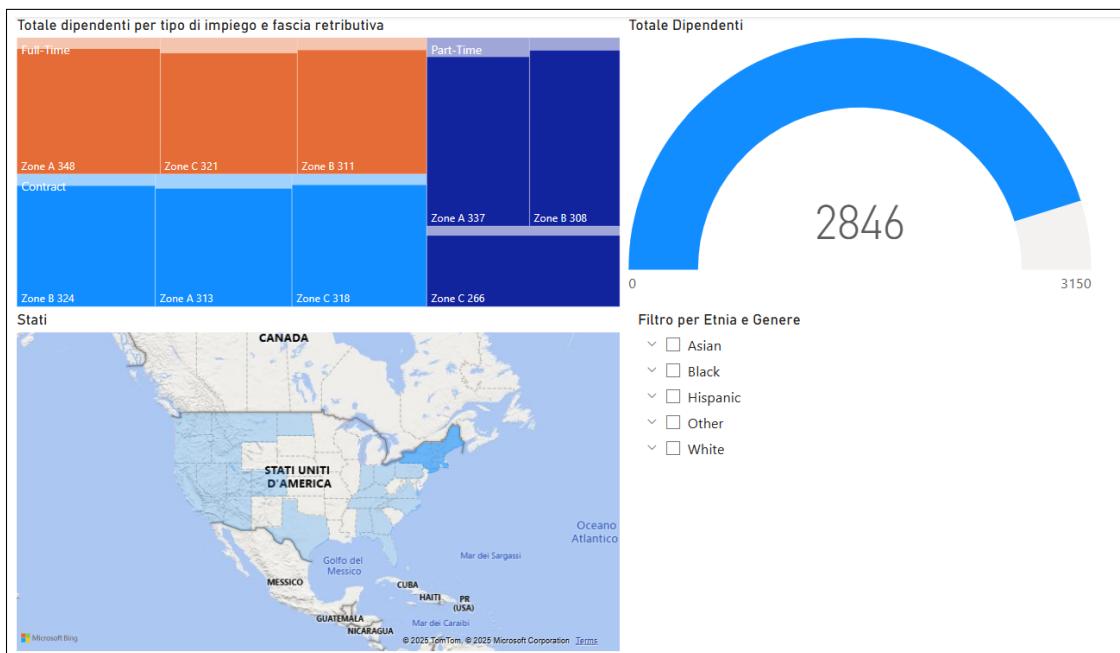


Figura 4.10: Distribuzione dei dipendenti per tipo di impiego e fascia retributiva in base allo stato.

4.6 Foglio 4: Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione

Il quarto foglio della nostra analisi in Power BI fornisce una visione dettagliata della distribuzione dei dipendenti in base al supervisore di riferimento e del loro livello di soddisfazione. Grazie a grafici interattivi e filtri dinamici, è possibile analizzare la relazione tra leadership e benessere dei dipendenti, individuando eventuali aree di miglioramento.

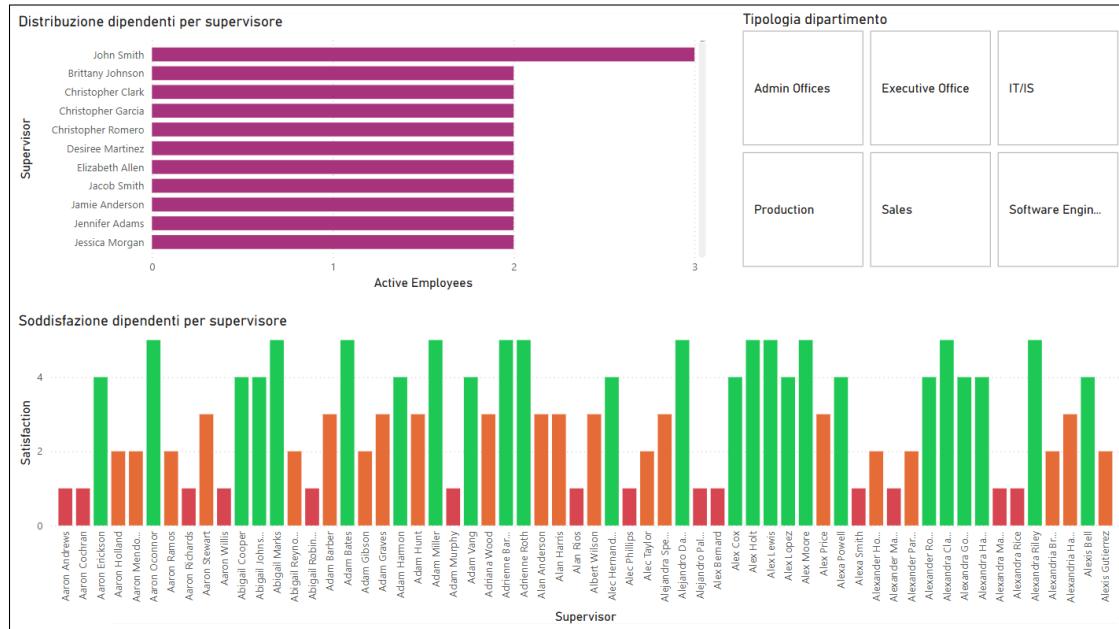


Figura 4.11: Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione.

Grazie all'integrazione dei diversi grafici e strumenti di filtro, questo foglio consente di:

- Identificare eventuali squilibri nella distribuzione dei dipendenti tra i vari supervisori.
- Valutare il grado di soddisfazione medio dei dipendenti per ogni supervisore e confrontarlo con la distribuzione del personale.
- Analizzare eventuali tendenze o differenze tra i dipartimenti aziendali per individuare aree di miglioramento nella gestione del personale.

4.6.1 Analisi

Il primo grafico, un istogramma a barre orizzontali, mostra la distribuzione dei dipendenti per ciascun supervisore. Questo consente di identificare quanti lavoratori fanno capo a ogni Figura di riferimento, permettendo un'analisi della struttura gerarchica aziendale. L'analisi ci permette di verificare se il carico di lavoro è distribuito equamente tra i supervisori. Dal grafico emerge che, ad eccezione di John Smith, che gestisce 3 dipendenti, tutti gli altri supervisori hanno ciascuno 2 dipendenti da gestire, indicando una distribuzione relativamente equilibrata.

Sotto, un secondo grafico a colonne verticali evidenzia il grado di soddisfazione medio dei dipendenti per ogni supervisore. Questo indicatore è fondamentale per valutare l'impatto della leadership sulla qualità dell'ambiente di lavoro e individuare eventuali criticità o best practice nella gestione del personale. Osserviamo che, nella maggior parte dei casi, il grado di soddisfazione medio è superiore a 3, mentre solo in alcune situazioni scende fino ad 1.

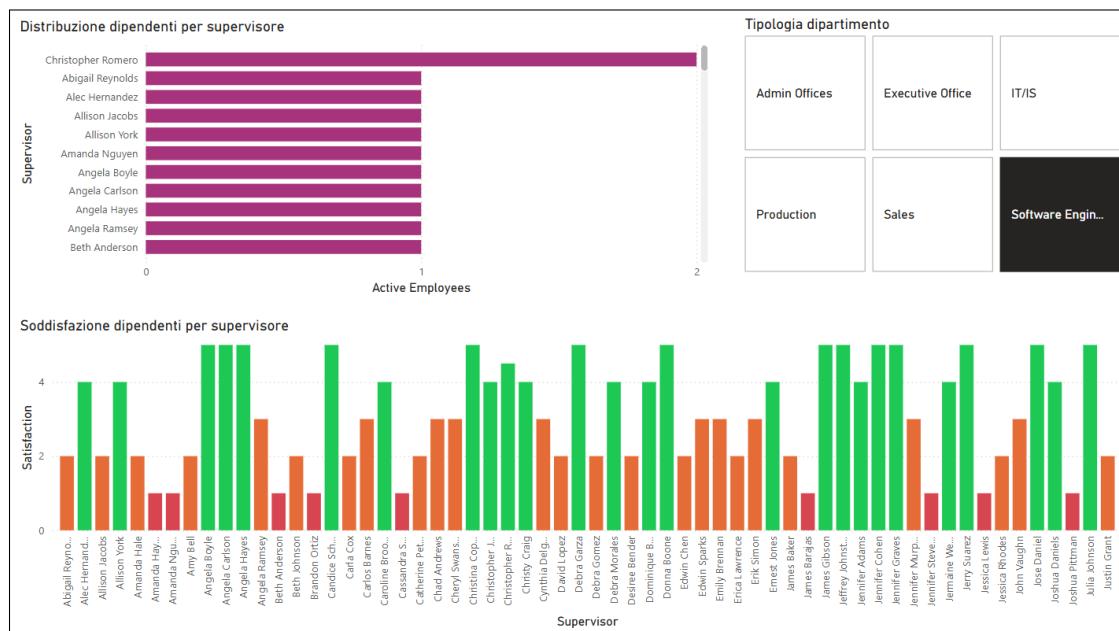


Figura 4.12: Analisi della distribuzione dei dipendenti in base al supervisore e del loro livello di soddisfazione per il dipartimento Software.

Un ulteriore elemento chiave dell'analisi è la possibilità di applicare filtri basati sul dipartimento come mostrato in Figura 4.12. Questo consente di focalizzare l'attenzione su specifiche aree aziendali e comprendere meglio eventuali discrepanze nella distribuzione del personale e nel livello di soddisfazione tra i diversi settori.

CAPITOLO 5

Conclusioni

5.1 Conclusioni

L'uso di strumenti di Business Intelligence si è rivelato estremamente utile per esplorare e visualizzare i dati aziendali in modo interattivo ed efficace. Tuttavia, durante il processo di analisi, abbiamo riscontrato alcune difficoltà legate alla qualità e alla struttura dei dataset disponibili.

Uno dei principali ostacoli è stata la difficoltà nel reperire dataset realmente consistenti e ricchi di informazioni significative. Sebbene molti dataset contengano un numero elevato di righe, spesso presentano un numero limitato di attributi descrittivi, riducendo di fatto le possibilità di condurre analisi approfondite. La mancanza di variabili di contesto o di informazioni più dettagliate sui dipendenti ha limitato la capacità di individuare relazioni significative tra i dati e di trarre insight realmente utili per il processo decisionale.

Un'altra problematica emersa durante l'analisi riguarda la qualità dei dati stessi. Nel nostro caso specifico, abbiamo notato che, sebbene gli attributi presenti nel dataset fossero promettenti per un'analisi dettagliata, i valori all'interno delle variabili sembravano essere generati artificialmente. Dopo un'osservazione più approfondita, è emerso che i dati erano probabilmente stati creati mediante un algoritmo che assegnava valori casuali a ciascun attributo, attingendo da un insieme predefinito di scelte.

Questa caratteristica ha avuto un impatto significativo sui risultati delle analisi. In particolare:

- Le previsioni basate sui dati storici tendevano a generare curve piatte e poco significative, in quanto non vi erano reali variazioni o trend evolutivi nel tempo, ma al contrario rappresentavano curve e oscillazioni marcate con comportamento

a denti di sega (in quanto i valori rappresentavano i margini superiori e inferiori del pool di scelta e tutti i valori intermedi).

- Le distribuzioni dei dati risultavano insolitamente omogenee, senza particolari squilibri o picchi in alcune categorie, suggerendo una mancanza di pattern reali.
- L'assenza di correlazioni significative tra le variabili ha reso difficile individuare insight concreti, poiché i dati non riflettevano dinamiche reali, ma piuttosto una distribuzione artificiale.

Inoltre, un altro punto molto sfavorevole, sono le sequenze temporali. Non avendo un dataset aggiornato, molte analisi considerate al giorno odierno, riportano misure superiori alle aspettative (vedi la durata media di un dipendente in azienda). Per questo motivo, seppur le analisi siano state pensate e realizzate correttamente, ancora una volta la mancanza di un dataset adeguato ci ha complicato il lavoro. Questi fattori hanno evidenziato l'importanza della qualità del dataset nella Business Intelligence. Strumenti avanzati possono fornire potenti capacità di analisi, ma la loro efficacia dipende fortemente dalla validità e dall'affidabilità dei dati di partenza. In futuro, per garantire analisi più significative e accurate, sarà fondamentale lavorare con dataset costruiti su dati reali e strutturati in modo da garantire una buona varietà di attributi e valori.

Nonostante queste difficoltà, l'analisi ha comunque fornito un'esperienza preziosa nell'utilizzo degli strumenti di BI e ha permesso di comprendere l'importanza di una corretta selezione e preparazione dei dati prima di eseguire qualsiasi tipo di elaborazione o previsione.