

Relazione Basi di Dati 2020-2021

Alessandro Zanatta, Christian Abbondo, Samuele Anzolin e Fabiana Rapticavoli

10-09-2020

1 Introduzione

Lo scopo del progetto è la progettazione e implementazione di una base di dati, a partire dall'analisi dei requisiti fino ad arrivare all'analisi dei dati.

La struttura del progetto è la seguente:

- **database/** - implementazione e analisi del database
 - **analysis/** - analisi in R
 - * **analysis.Rmd**
 - **Dockerfile** - **docker-compose.yml** - dockerfiles per il database
 - **scripts/** - script per il popolamento del database
 - **sql/** - file di implementazione del database
 - **Makefile** - utilizzato per semplificare l'uso del database

2 Analisi dei requisiti

Iniziamo riportando il testo del progetto assegnato:

Si vuole realizzare una base di dati per la gestione giornaliera di un sistema di trasporto urbano mediante autobus caratterizzato dal seguente insieme di requisiti:

- Il servizio sia organizzato in un certo insieme di linee di trasporto urbano. Ogni linea sia identificata univocamente da un numero (ad esempio, linea di trasporto urbano numero 1) e sia caratterizzata da un certo numero di fermate. Si tenga traccia anche dell'ordine della fermate di una data linea.
- Ogni fermata sia identificata univocamente da un nome. Ad ogni fermata sia, inoltre, associato un indirizzo. Una fermata possa essere raggiunta da più linee di trasporto.
- Ogni linea è servita, nello stesso giorno, da uno o più autobus. Ogni autobus è identificato univocamente dal numero di immatricolazione ed è caratterizzato da un numero di posti (in piedi e a sedere). Un autobus può essere guidato da guidatori diversi. Si vuol tener traccia dell'autobus e del guidatore che ha eseguito una corsa (su una certa linea, una corsa corrisponde ad un'istanza di linea effettuata ad una certa ora e giorno).
- L'autista di un autobus è identificabile univocamente tramite il suo codice fiscale o per mezzo del numero di patente ed è caratterizzato da nome, cognome, data di nascita e luogo di residenza e da zero o più numeri di recapito telefonico.
- I clienti che usufruiscono del servizio di trasporto possono viaggiare tramite biglietti giornalieri o tramite abbonamento. Si è interessati unicamente alla memorizzazione degli utenti tesserati. Si vuole, in particolare, tener traccia dei loro spostamenti, cioè si vogliono conoscere le corse di cui ha usufruito e il numero di corse utilizzate mensilmente.
- Un abbonamento è identificato univocamente da un numero di tessera e dalla data di inizio (si suppone che ad una tessera non siano collegati abbonamenti con la stessa data di inizio in quanto un singolo abbonamento permette di usufruire di qualsiasi autobus), è caratterizzato da un tipo di abbonamento (mensile, trimestrale, annuale). Ogni tessera è intestata a uno e un solo cliente il quale è, a sua volta, identificato tramite codice fiscale ed è caratterizzato da nome, cognome, data di nascita e luogo di residenza e da un recapito telefonico (utilizzato in caso di smarrimento dell'abbonamento stesso). Si vuole tener traccia di tutti gli abbonamenti sottoscritti dal cliente tesserato.

Il **dominio di interesse** della base di dati è definito nella prima frase ed è il *sistema di trasporto urbano mediante autobus*.

2.1 Glossario dei termini

Termine	Descrizione	Sinonimi	Collegamenti
<i>Sistema di trasporto urbano</i>	Dominio della base di dati	Servizio	
<i>Linea di trasporto urbana</i>	Successione di fermate dell'autobus	Linea di trasporto, Linea	Fermata, Autobus, Corsa
<i>Fermata</i>	Punto di salita e/o discesa di un determinato luogo		Linea di trasporto urbana
<i>Autobus</i>	Mezzo utilizzato per il trasporto dei clienti		Linea di trasporto urbana, Autista, Cliente, Corsa
<i>Autista</i>	Conducente degli autobus	Guidatore	Autobus, Corsa
<i>Corsa</i>	Istanza di una linea, eseguita da un certo autista su un certo autobus in una data ora e giornata		Autobus, Autista, Linea di trasporto urbana, Cliente
<i>Cliente</i>	Usufruttore del sistema di trasporto urbano	Utente tesserato, Cliente tesserato	Autobus, Corsa, Tessera
<i>Tessera</i>	Documento associato ad un singolo cliente		Cliente, Abbonamento
<i>Abbonamento</i>	Contratto fra l'azienda offrente il sistema di trasporto urbano e il cliente		Tessera

2.2 Strutturazione dei requisiti

Riscriviamo ora quindi i requisiti ristrutturandoli in gruppi di frasi omogenee, relative cioè agli stessi concetti.

Frasi di *carattere generale*

Si vuole realizzare una base di dati per la gestione giornaliera di un sistema di trasporto urbano mediante autobus.

Frasi relative alle *linee di trasporto urbane*

Il sistema di trasporto urbano sia organizzato in un certo insieme di linee di trasporto urbano. Ogni linea di trasporto urbana sia identificata univocamente da un numero e sia caratterizzata da un certo numero di fermate. Si tenga traccia anche dell'ordine delle fermate di una data linea di trasporto urbana. Ogni linea di trasporto urbana è servita, nello stesso giorno, da uno o più autobus.

Frasi relative alle *fermate*

Ogni fermata sia identificata univocamente da un nome. Ad ogni fermata sia, inoltre, associato un indirizzo. Una fermata possa essere raggiunta da più linee di trasporto urbane.

Frase relative agli ***autobus***

Ogni autobus è identificato univocamente dal numero di immatricolazione (targa) ed è caratterizzato da un numero di posti (in piedi e a sedere). Un autobus può essere guidato da autisti diversi.

Frase relative agli ***autisti***

L'autista di un autobus è identificabile univocamente tramite il suo codice fiscale o per mezzo del numero di patente ed è caratterizzato da nome, cognome, data di nascita e luogo di residenza e da zero o più numeri di recapito telefonico.

Frase relative alle ***corse***

Si vuole tener traccia dell'autobus e dell'autista che ha eseguito una corsa, corrispondente cioè ad una istanza di linea di trasporto urbana effettuata ad una certa ora e giorno

Frase relative ai ***clienti***

Il cliente è identificato dal codice fiscale ed è caratterizzato da nome, cognome, data di nascita e luogo di residenza e da un recapito telefonico (in caso di smarrimento dell'abbonamento). Si è interessati alla memorizzazione dei clienti possessori di tessera. Si vuole, in particolare, tener traccia dei loro spostamenti, cioè si vogliono conoscere le corse di cui ha usufruito, e del numero di corse utilizzate mensilmente.

Frase relative alle ***tessere***

Ogni tessera è intestata ad uno e un solo cliente.

Frase relative agli ***abbonamenti***

Un abbonamento è identificato univocamente da un numero di tessera e dalla data di inizio (si suppone che ad una tessera non siano collegati abbonamenti con la stessa data di inizio in quanto un singolo abbonamento permette di usufruire di qualsiasi autobus), è caratterizzato da un tipo di abbonamento (mensile, trimestrale, annuale). Si vuole tener traccia di tutti gli abbonamenti sottoscritti dal cliente tesserato.

2.3 Operazioni frequenti sui dati

1. Inserimento di una nuova fermata in una linea di trasporto (15 all'anno)
2. Inserimento di una nuova corsa (450 al giorno)
3. Inserimento di una nuova corsa effettuata da un cliente (12500 al giorno)
4. Lettura del numero di corse mensili effettuate (1 al mese)
5. Inserimento di un nuovo cliente (50 al mese)
6. Inserimento di un nuovo abbonamento (2500 al mese)
7. Stampa del numero di fermate per una data linea di trasporto urbana (una alla settimana)
8. Stampa del numero di abbonamenti attualmente validi (una al giorno)

3 Progettazione concettuale

3.1 Analisi preliminare

Iniziamo preliminarmente da un'analisi del testo di specifiche dei requisiti utilizzando la seguente legenda:

Legenda	Entità	Attributo	Relazioni	Attributi di relazione
---------	--------	-----------	-----------	------------------------

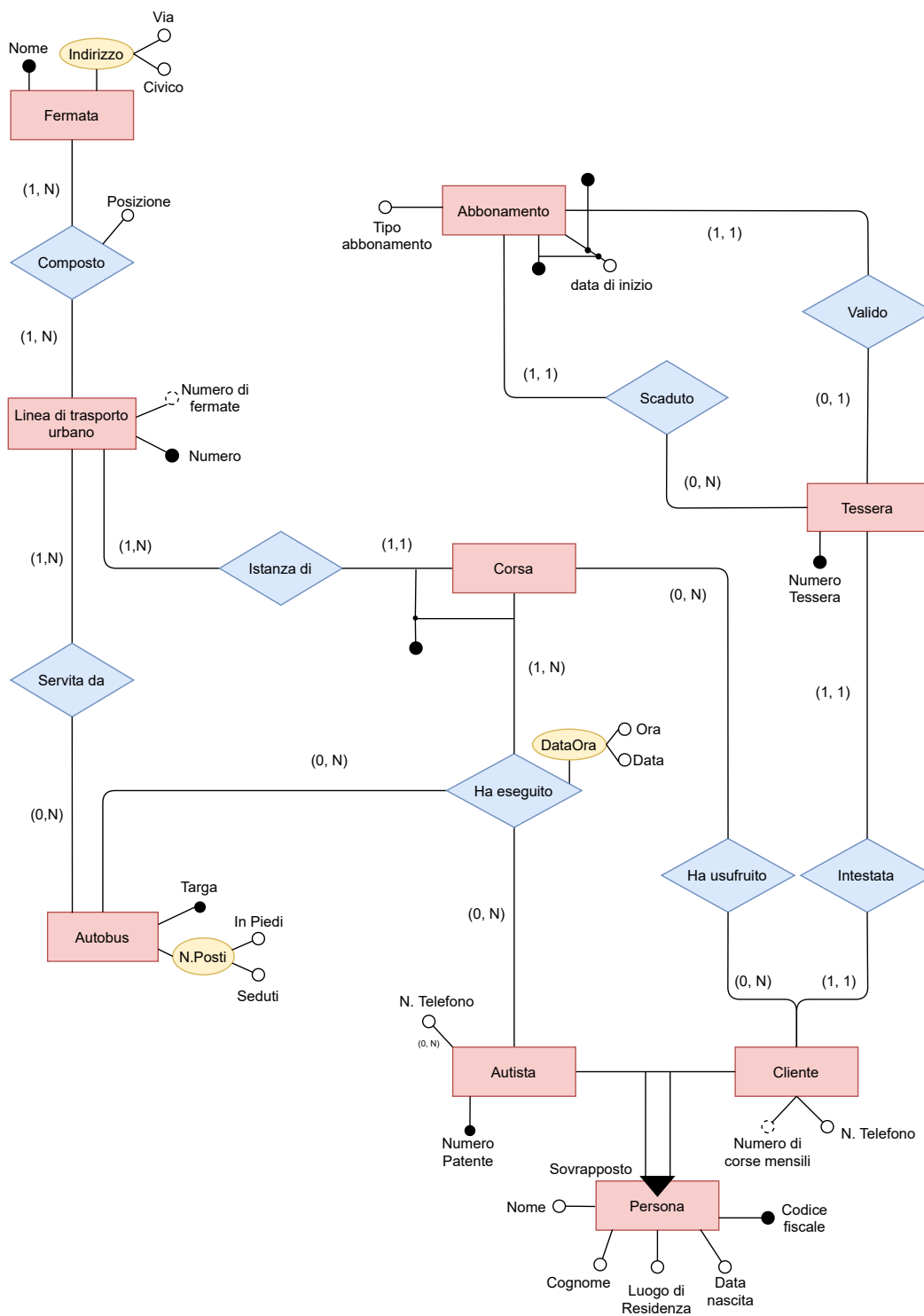
Si vuole realizzare una base di dati per la gestione giornaliera di un sistema di trasporto urbano mediante autobus:

- Il sistema di trasporto urbano sia organizzato in un certo insieme di **linee di trasporto urbano**. Ogni linea di trasporto urbana sia identificata univocamente da un **numero** e sia caratterizzata da un certo **numero di fermate**. Si tenga traccia anche dell'**ordine** delle **fermate** di una data linea di trasporto urbana. Ogni linea di trasporto urbana **è servita**, nello stesso giorno, da uno o più **autobus**.
- Ogni fermata sia identificata univocamente da un **nome**. Ad ogni fermata sia, inoltre, associato un **indirizzo**. Una fermata possa essere raggiunta da più linee di trasporto urbane.
- Ogni autobus è identificato univocamente dal **numero di immatricolazione (targa)** ed è caratterizzato da un **numero di posti (in piedi e a sedere)**. Un autobus può **essere guidato da autisti** diversi.
- L'autista di un autobus è identificabile univocamente tramite il suo **codice fiscale** o per mezzo del **numero di patente** ed è caratterizzato da **nome, cognome, data di nascita e luogo di residenza** e da zero o più **numeri di recapito telefonico**.
- Si vuole tener traccia dell'autobus e dell'autista che **ha eseguito** una corsa, corrispondente cioè ad una istanza di linea di trasporto urbana effettuata ad una certa **ora** e **giorno**.
- Il **cliente** è identificato dal **codice fiscale** ed è caratterizzato da **nome, cognome, data di nascita e luogo di residenza** e da un **recapito telefonico** (in caso di smarrimento dell'abbonamento). Si è interessati alla memorizzazione dei clienti possessori di **tessera**. Si vuole, in particolare, tener traccia dei loro spostamenti, cioè si vogliono conoscere le corse di cui **ha usufruito**, e del **numero di corse utilizzate mensilmente**.
- Ogni tessera è **intestata a** uno e un solo cliente.
- Un **abbonamento** è identificato univocamente da un **numero di tessera** e dalla **data di inizio** (si suppone che ad una tessera non siano collegati abbonamenti con la stessa data di inizio in quanto un singolo abbonamento permette di usufruire di qualsiasi autobus), è caratterizzato da un **tipo di abbonamento** (mensile, trimestrale, annuale). Si vuole tener traccia di tutti gli abbonamenti sottoscritti dal cliente tesserato.

3.2 Schema Entità-Relazioni

La strategia di progetto utilizzata è la strategia mista.

Si riporta di seguito il diagramma Entità-Relazioni risultante dallo sviluppo delle specifiche:



3.2.1 Vincoli aggiuntivi

Si evidenziano i seguenti *vincoli aziendali*:

- Nella relazione **composto**, per la stessa istanza di **linea di trasporto urbana**, non sono ammessi attributi **posizione** duplicati. Si noti che l'attributo **posizione** è inserito nello schema al fine di tener conto dell'ordine delle **fermate** della singola **linea di trasporto urbana**.
- Ogni **tessera** deve essere in relazione con almeno una delle due relazioni **valido** e **scaduto**. Si assume infatti che ogni cliente abbia almeno un abbonamento, sia questo scaduto o ancora in vigore.

Si indicano inoltre le seguenti *regole di derivazione*:

- L'attributo **numero di fermate** dell'entità **linea di trasporto urbana** è derivabile contando il numero di istanze di **fermata** in relazione con una data istanza di **linea di trasporto urbana**.
- L'attributo **numero di corse mensili** dell'entità **cliente** è derivabile contando il numero di istanze di **corsa** in un dato mese in relazione con una data istanza di **cliente**.

3.2.2 Considerazioni

Nello schema sono presenti dei cicli. Dato che i cicli potrebbero essere, potenzialmente, fonti di incoerenze, segue una loro analisi:

1. Linea di trasporto - autobus - corsa
 - Questo ciclo è problematico. Un **autobus** non può, infatti, essere utilizzato per eseguire una **corsa** che è istanza di una **linea di trasporto urbana** diversa da quella servita dall'**autobus** stesso.
2. Tessera - abbonamento
 - Anche questo ciclo è problematico. Risulta infatti insensato che un **abbonamento** possa essere in relazione con una **tessera** sia in qualità di **abbonamento** valido che in qualità di **abbonamento** scaduto.

Si riportano alcune decisioni progettuali attinenti alle cardinalità delle relazioni:

- Il vincolo di partecipazione opzionale è stato utilizzato, nella relazione ternaria **ha eseguito**, sia per **autobus** che per **autista** in quanto:
 - Un **autobus** potrebbe, se esso è nuovo o di riserva, non aver effettuato alcuna **corsa**.
 - Un **autista** potrebbe, nel caso in cui questo fosse neo-assunto, non aver eseguito alcuna **corsa**.
- Per considerazioni analoghe a quelle precedenti, la partecipazione di **autobus** alla relazione **servita da** è opzionale. Un **autobus** potrebbe, infatti, non servire alcuna **linea di trasporto urbano**.
- La partecipazione nella relazione **ha usufruito** è, da ambo i lati, opzionale in quanto:
 - Una **corsa** potrebbe non avere alcun **cliente** tra i passeggeri del mezzo.
 - Un **cliente** potrebbe, se appena tesserato, non aver usufruito di alcuna **corsa**.

4 Progettazione logica

4.1 Ristrutturazione dello schema Entità Relazioni

4.1.1 Analisi delle ridondanze

Si definiscono, innanzitutto, i volumi dei dati e si riportano le frequenze delle operazioni:

Concetto	Tipo	Volume
Fermata	Entità	225
Linea di trasporto urbana	Entità	15
Autobus	Entità	100
Corsa	Entità	450 al giorno
Autista	Entità	75
Cliente	Entità	7500
Tessera	Entità	7500
Abbonamento	Entità	100000
Composto	Relazione	300
Servita da	Relazione	90
Istanza di	Relazione	450 al giorno
Ha eseguito	Relazione	450 al giorno
Ha usufruito	Relazione	12000
Intestata	Relazione	7500
Scaduto	Relazione	92500
Valido	Relazione	7500

Operazione	Frequenza
Inserimento di una nuova fermata in una linea di trasporto	15 all'anno
Inserimento di una nuova corsa	450 al giorno
Inserimento di una nuova corsa effettuata da un cliente	12500 al giorno
Lettura del numero di corse mensili effettuate	1 al mese
Inserimento di un nuovo cliente	50 al mese
Inserimento di un nuovo abbonamento	2500 al mese
Stampa del numero di fermate per una data linea di trasporto urbana	1 alla settimana
Stampa del numero di abbonamenti attualmente validi	1 al giorno

In base alle tabelle sopra definite, si è proseguito a creare le tavole degli accessi relative alle operazioni che presentano ridondanze.

Numero di fermate Segue l'analisi di vantaggi e svantaggi dell'attributo derivato **numero di fermate**, il quale riguarda le seguenti interrogazioni:

1. Inserimento di una nuova fermata in una linea di trasporto

Concetto	Costrutto	Accessi	Tipo
Fermata	Entità	1	Scrittura
Composto	Relazione	1	Scrittura
Linea di trasporto urbano	Entità	1	Lettura
Linea di trasporto urbano	Entità	1	Scrittura

Tabella 10: Tabella degli accessi in presenza di ridondanza

Concetto	Costrutto	Accessi	Tipo
Fermata	Entità	1	Scrittura
Composto	Relazione	1	Scrittura

Tabella 11: Tabella degli accessi senza ridondanze

2. Stampa del numero di fermate per una data linea di trasporto urbana

Concetto	Costrutto	Accessi	Tipo
Linea di trasporto urbano	Entità	1	Lettura

Tabella 12: Tabella degli accessi in presenza di ridondanza

Concetto	Costrutto	Accessi	Tipo
Linea di trasporto urbano	Entità	1	Lettura
Composto	Relazione	20	Lettura

Tabella 13: Tabella degli accessi senza ridondanze

Gli accessi, in lettura e scrittura, all'entità **linea di trasporto urbano**, richiesto in caso di ridondanza, sono necessari per garantire che l'attributo derivato e ridondante **numero di fermate** sia aggiornato correttamente.

I seguenti calcoli sono effettuati dando un peso doppio agli accessi in scrittura dato che questi risultano essere più dispendiosi.

Il numero di accessi totale per la prima operazione è di $(2 + 2 + 1 + 2) \cdot 15 = 105$ accessi all'anno in caso di ridondanza. Il numero di accessi annui, in caso di assenza di dato ridondante, è di $(2 + 2) \cdot 15 = 60$.

Il numero di accessi totale per la seconda operazione è di $1 \cdot \lfloor \frac{365}{7} \rfloor = 52$ accessi all'anno in caso di ridondanza. In caso di assenza, il numero degli accessi sale a $21 \cdot \lfloor \frac{365}{7} \rfloor = 1092$ all'anno.

Si può quindi notare che, dato il maggior numero di accessi necessari in caso di mancanza della ridondanza (1152 contro 157), sia conveniente mantenere l'attributo derivato **numero di fermate**, a costo di un piccolo utilizzo aggiuntivo di memoria.

Numero di corse mensili Segue, infine, l'analisi di vantaggi e svantaggi dell'attributo derivato **numero di corse mensili**, il quale riguarda le seguenti interrogazioni:

1. Inserimento di una nuova corsa effettuata da un cliente

Concetto	Costrutto	Accessi	Tipo
Ha usufruito	Relazione	1	Scrittura
Cliente	Entità	1	Lettura
Cliente	Entità	1	Scrittura

Tabella 14: Tabella degli accessi in presenza di ridondanza

Concetto	Costrutto	Accessi	Tipo
Ha usufruito	Relazione	1	Scrittura

Tabella 15: Tabella degli accessi senza ridondanze

2. Lettura del numero di corse mensili effettuate¹

Concetto	Costrutto	Accessi	Tipo
Cliente	Entità	1	Lettura

Tabella 16: Tabella degli accessi in presenza di ridondanza

Concetto	Costrutto	Accessi	Tipo
Cliente	Entità	1	Lettura
Composto	Relazione	45	Lettura
Corsa	Entità	45	Lettura

Tabella 17: Tabella degli accessi senza ridondanze

L'accesso, in lettura e scrittura, all'entità **cliente**, richiesto in caso di ridondanza, è necessario per garantire che l'attributo derivato e ridondante **numero di corse mensili** sia aggiornato correttamente.

Si ricorda che i seguenti calcoli sono effettuati dando un peso doppio agli accessi in scrittura.

Il numero di accessi totale per la prima operazione è di $(2 + 1 + 2) \cdot (12000 \cdot 28) = 1680000$ accessi al mese in caso di ridondanza, dove 28 è la durata di una mensilità. Il numero di accessi mensili, in caso di assenza di dato ridondante, è di $2 \cdot (12000 \cdot 28) = 672000$.

Il numero di accessi totale per la seconda operazione è di $1 \cdot 7500 = 7500$ accessi al mese in caso di ridondanza, dove 7500 è il numero di istanze attese dell'entità **cliente**. In caso di assenza, il numero degli accessi è invece di $(1 + 45 + 45) \cdot 7500 = 682500$ al mese.

Ricapitolando, si hanno:

- 1687500 accessi in caso di ridondanza
- 1354500 accessi in assenza di ridondanza

Risulta quindi conveniente **non** mantenere l'attributo **numero di corse mensili**.

¹La tabella si riferisce agli accessi necessari per ogni cliente. I calcoli relativi alla totalità sono effettuati in seguito.

4.1.2 Eliminazione delle generalizzazioni

L'unica generalizzazione presente nello schema E-R è quella che riguarda le entità **persona**, **autista** e **cliente**. Fra le tre possibili alternative di rimozione della generalizzazione, si è scelto di accorpare nei figli (**autista** e **cliente**) l'entità genitore (**persona**). Si è deciso di procedere in tale maniera in quanto risulta maggiormente conveniente in termini sia di memoria che di numero di accessi richiesti in quanto le interrogazioni più frequenti individuate riguardano le entità figlie.

4.1.3 Partizionamento/accorpamento di concetti

Eliminazione degli attributi composti Gli attributi composti sono stati eliminati nel seguente modo:

- Per quanto riguarda l'attributo **indirizzo** relativo a **fermata**, è stato scelto di eliminare gli attributi figli in quanto si è maggiormente interessati all'indirizzo nella sua interezza;
- Considerando l'attributo **dataora** dell'entità **ha eseguito**, si è deciso di mantenere l'attributo genitore, in quanto si prevede di accedervi nella sua interezza;
- In merito all'attributo **n. posti** dell'entità **autobus**, infine, è stato scomposto l'attributo dato che questo sarebbe risultato difficilmente rappresentabile unitamente.

Partizionamento di entità Dato che non è stato ritenuto necessario alcun partizionamento, lo schema E-R non ha subito modifiche in tal senso.

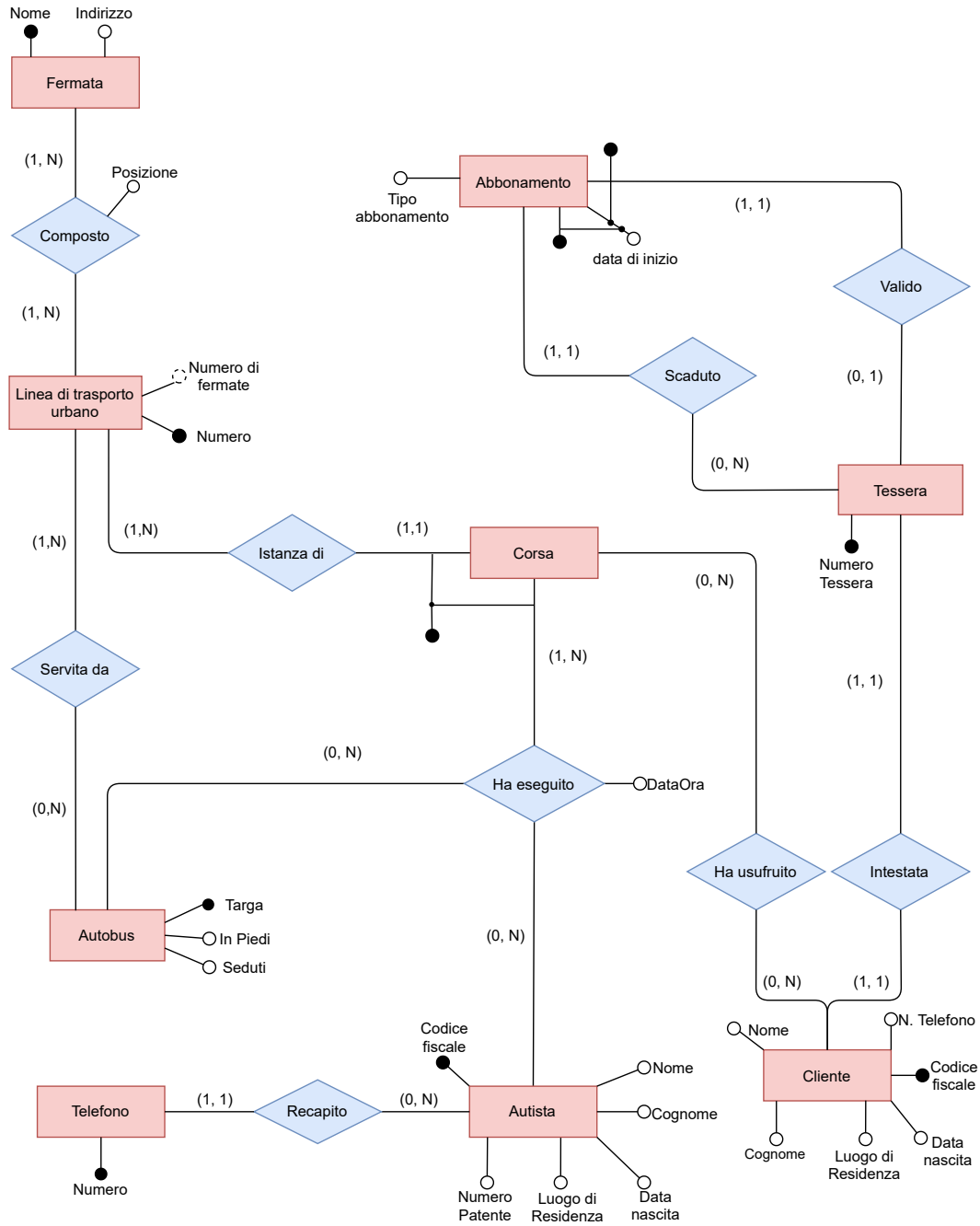
Eliminazione di attributi multivalore L'unico attributo multivalore presente riguarda i **recapiti telefonici** relativi all'entità **autista**. Si è provveduto a partizionarlo in due entità, **autista** e **telefono**, legate da una relazione uno a molti (opzionale dal lato dell'**autista**).

Accorpamento di entità Non si è ritenuto necessario accorpare alcuna entità anche data la scarsità di relazioni di tipo uno a uno.

4.1.4 Scelta degli identificatori principali

L'unica entità con due chiavi candidate è l'entità **autista**. Al fine di avere una maggiore coerenza con la chiave utilizzata dall'entità **cliente**, si è scelto di mantenere come identificatore principale l'attributo **codice fiscale**.

Si riporta nella pagina seguente, quindi, il diagramma E-R ristrutturato secondo le considerazioni sopra elencate.



4.2 Traduzione verso il modello relazionale

In riferimento all'ultimo schema ER, si è quindi proceduto alla stesura dello schema relazionale. Si noti che, durante la traduzione, sono state eseguite alcune rinomine, al fine di rendere lo schema più intuitivo. Sono inoltre stati introdotti alcuni vincoli aggiuntivi sui valori che alcuni attributi possono assumere, dei quali non si era tenuto conto nelle precedenti fasi di progettazione.

Si divide la traduzione in base alle tipologie di concetto da tradurre. In particolare si evidenziano:

- Entità
- Relazioni molti a molti
- Relazioni uno a molti
- Relazioni uno a uno

4.2.1 Entità

Le seguenti entità vengono di seguito tradotte in modo definitivo:

- Fermata
- Linea di trasporto urbano
- Autobus
- Autista
- Corsa
- Tesserata

Le entità **cliente**, **abbonamento**, **tesserata** e **telefono** verranno invece trattate anche durante la traduzione di alcune relazioni a cui sono legate.

Traduzioni definitive

- **Fermata**(Nome, Indirizzo)
 - Vincolo NotNull: Indirizzo
- **LineaTrasportoUrbano**(Numero, NumeroFermate)
 - Vincolo NotNull: NumeroFermate
 - Vincoli aggiuntivi: NumeroFermate > 0
- **Autobus**(Targa, InPiedi, Seduti)
 - Vincolo NotNull: InPiedi, Seduti
 - Vincoli aggiuntivi: InPiedi + Seduti > 0
- **Autista**(CodiceFiscale, Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroPatente)
 - Vincolo NotNull: Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroPatente
 - Vincolo Unique: NumeroPatente
- **Corsa**(DataOra, NumeroLinea)
 - Vincolo di chiave esterna: NumeroLinea si riferisce alla chiave primaria di **LineaTrasportoUrbano**, DataOra si riferisce alla colonna **DataOra** di **HaEseguito**²

Traduzioni che subiranno modifiche

- **Cliente**(CodiceFiscale, Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroTelefono)
 - Vincolo NotNull: Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroTelefono
- **Abbonamento**(DataInizio, Tesserata, TipoAbbonamento)
 - Vincolo NotNull: TipoAbbonamento
 - Vincolo di chiave esterna: Tesserata si riferisce alla chiave primaria di **Tesserata**
- **Tesserata**(NumeroTesserata)
- **Telefono**(Numero)

4.2.2 Relazioni molti a molti

Questi tipi di relazioni vengono sempre tradotte introducendo una nuova relazione nello schema relazionale.

²Questa relazione è definita nella sezione 4.2.2

È importante notare che non tutti i vincoli di questo tipo di relazioni sono direttamente traducibili in SQL senza perdita di informazione, come verrà meglio indicato in seguito.

- **Composto**(NomeFermata, NumeroLinea, Posizione)
 - Vincolo NotNull: Posizione
 - Vincolo di chiave esterna: NomeFermata fa riferimento alla colonna Nome della relazione Fermata, NumeroLinea fa riferimento alla colonna Numero della tabella LineaTrasportoUrbano
 - Vincolo unique: la tripla (NomeFermata, NumeroLinea, Posizione) deve essere univoca per evitare situazioni invalide in cui diverse fermate della stessa linea si trovano nella stessa posizione
 - Vincoli aggiuntivi: Posizione > 0
- **ServitaDa**(NumeroLinea, Targa)
 - Vincolo di chiave esterna: NumeroLinea fa riferimento alla colonna Numero della tabella LineaTrasportoUrbano, Targa fa riferimento alla colonna Targa della tabella Autobus
- **HaEseguito**(Autobus, Autista, DataOra)
 - Vincolo di chiave esterna: Autobus fa riferimento alla colonna Targa della tabella Autobus, Autista fa riferimento alla colonna CodiceFiscale della tabella Autista
 - Vincoli aggiuntivi: DataOra non può riferirsi al futuro
- **HaUsufruito**(Cliente, DataOra, NumeroLinea)
 - Vincolo di chiave esterna: Cliente fa riferimento alla colonna CodiceFiscale della tabella Cliente, DataOra fa riferimento alla colonna DataOra della tabella Corsa, NumeroLinea fa riferimento alla colonna NumeroLinea della tabella Corsa

Alcuni vincoli, come riportato sopra, non sono esprimibili in quanto la relazione di sottoinsieme, data da una chiave esterna, non è sufficiente a garantire la partecipazione obbligatoria di una data entità ad una certa relazione in caso di relazione molti a molti. Vengono, quindi, esplicitati alcuni vincoli aggiuntivi che verranno gestiti tramite appositi trigger:

- Vincolo di partecipazione obbligatoria fra l'entità **fermata** e la relazione **composto**;
- Vincolo di partecipazione obbligatoria fra l'entità **linea di trasporto urbano** e la relazione **composto**;
- Vincolo di partecipazione obbligatoria fra l'entità **linea di trasporto urbano** e la relazione **istanza di** (la quale è stata codificata tramite una chiave esterna in **corsa**);
- Vincolo di partecipazione obbligatoria fra l'entità **linea di trasporto urbano** e la relazione **servita da**;

4.2.3 Relazioni uno a molti

Questo tipo di relazioni è, solitamente, traducibile aggiungendo una chiave esterna nell'entità nel lato con cardinalità massima uno. Dato che l'entità **corsa** è debole, la traduzione della relazione **istanza di** è già stata effettuata in precedenza e non verrà trattata di seguito.

Recapito Viene apportata la seguente modifica alla relazione precedentemente definita:

- **Telefono**(Numero, Autista)
 - Vincolo NotNull: Autista
 - Vincolo di chiave esterna: Autista fa riferimento alla colonna CodiceFiscale della tabella Autista

Scaduto e Valido Queste due relazioni vengono trattate congiuntamente in quanto entrambe legano le entità **abbonamento** e **tessera**. Non è pertanto possibile aggiungere semplicemente una chiave esterna nella traduzione di **abbonamento** in quanto questo renderebbe impossibile codificare in SQL la differenza semantica delle due relazioni. La soluzione che si è trovata in prima battuta è quella di aggiungere un attributo booleano **AbbonamentoValido** in **abbonamento**. Si noti inoltre che questo attributo non sarebbe nemmeno necessario in quanto sarebbe sufficiente derivare la validità dell'abbonamento confrontando la data di inizio dell'abbonamento stesso e la sua tipologia (la quale indica il periodo di durata).

Una seconda alternativa sarebbe, invece, quella di mantenere la relazione **valido**.

Nonostante la prima soluzione potrebbe, in apparenza, risultare migliore, si può notare che non risulterebbe efficiente. L'operazione **Stampa del numero di abbonamenti attualmente validi**, infatti, sarebbe estremamente più dispendiosa in termini di numero di accessi necessari. Si forniscono di seguito le relative tabelle.

Concetto	Costrutto	Accessi	Tipo
Abbonamento	Entità	92500+7500	Lettura

Tabella 18: Tabella degli accessi della soluzione uno

Concetto	Costrutto	Accessi	Tipo
Valido	Relazione	7500	Lettura

Tabella 19: Tabella degli accessi della soluzione due

Si può quindi notare come il numero di accessi, utilizzando la seconda soluzione, sia inferiore di più di un ordine di grandezza. Le tabelle sono quindi definite nel seguente modo: ³

- **Valido**(Tessera, DataInizio)
 - Vincoli di chiave esterna: Tessera fa riferimento alla chiave primaria della tabella **Tessera**, DataInizio fa riferimento alla colonna **DataInizio** di **Abbonamento**
 - Vincolo unique: Tessera
- **Scaduto**(Tessera, DataInizio)
 - Vincoli di chiave esterna: Tessera fa riferimento alla chiave primaria della tabella **Tessera**, DataInizio fa riferimento alla colonna **DataInizio** di **Abbonamento**

4.2.4 Relazioni uno a uno

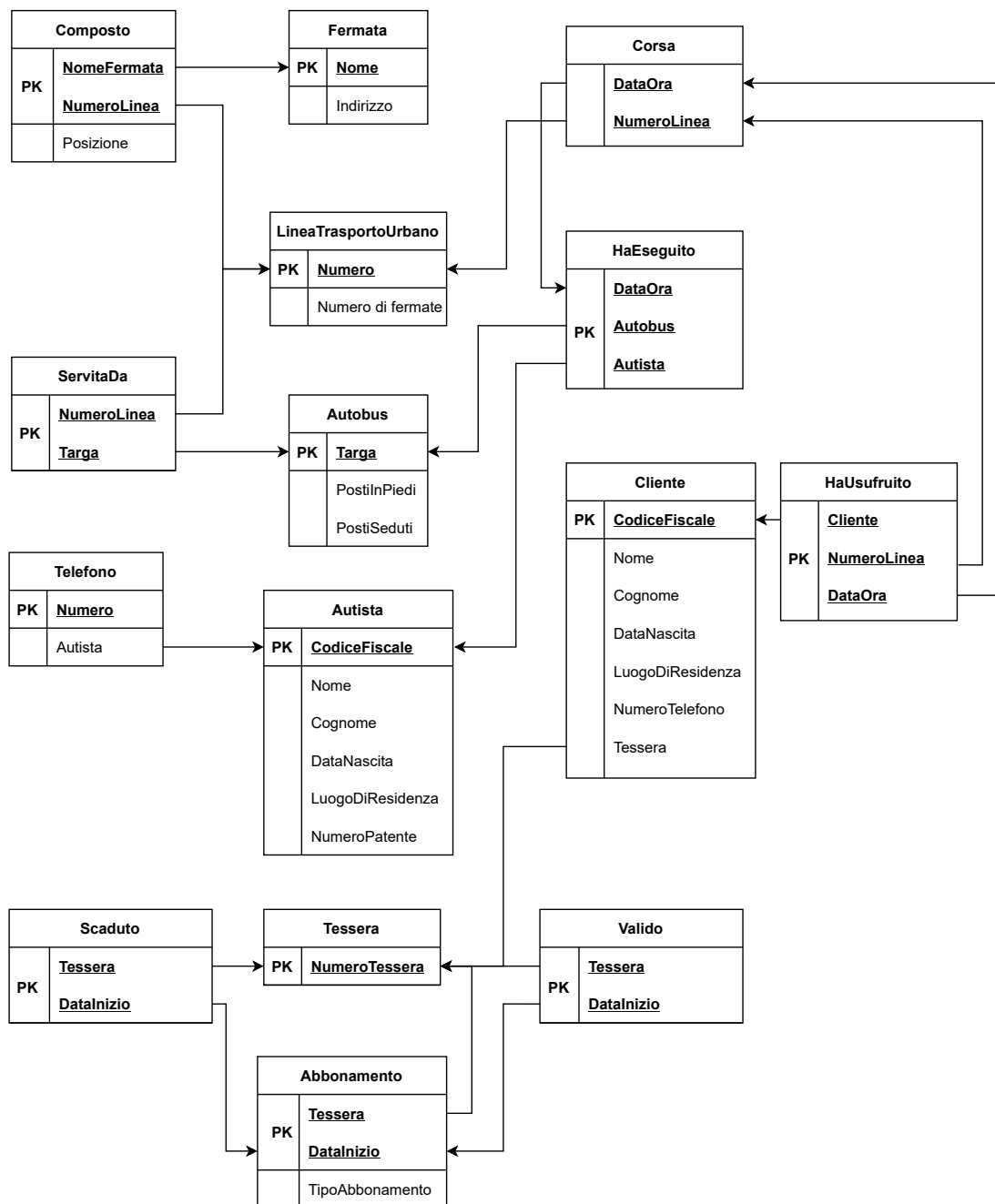
Vi è una unica relazione uno a uno, quella che lega **cliente** a **tessera** tramite la relazione **intestata**. In questo caso è sufficiente aggiungere un chiave esterna ad una ed una sola entità, a scelta fra le due in relazione. Non essendoci operazioni riguardanti le tessere, si è scelto di aggiungere la chiave esterna nella relazione **cliente**, in quanto questo risulta maggiormente naturale. La definizione della relazione **cliente** è stata quindi modificata come segue:

- **Cliente**(CodiceFiscale, Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroTelefono, Tessera)
 - Vincolo NotNull: Nome, Cognome, DataNascita, LuogoDiResidenza, NumeroTelefono, Tessera
 - Vincolo Unique: Tessera
 - Vincolo di chiave esterna: Tessera fa riferimento alla colonna **NumeroTessera** della tabella **Tessera**

³Si analizzeranno gli effetti di questa decisione in seguito, in particolare durante la fase di analisi dei dati.

4.2.5 Schema finale

Si riporta di seguito una visualizzazione dello schema.



5 Progettazione fisica

Per la fase di progettazione fisica si sono analizzate le [operazioni frequenti definite precedentemente](#).

Riguardo l'operazione **lettura del numero di corse mensili effettuate**, la creazione di un indice sull'attributo **DataOra** di **Corsa** potrebbe rendere il reperimento delle corse di interesse per l'interrogazione più efficiente.

Considerando le restanti interrogazioni, non si ritiene sia necessario aggiungere ulteriori indici. Si noti, per esempio, che la query **stampa del numero di fermate per una data linea di trasporto urbana** non beneficerebbe in alcun modo di un indice grazie all'attributo ridondante **NumeroFermate**.

Si noti che si utilizzeranno alcuni **strumenti** offerti dal DBMS postgres per analizzare l'efficacia dell'indice definito. Questo verrà effettuato una volta popolato il database completamente.

6 Implementazione

Il codice relativo all'implementazione è disponibile nella cartella `database/sql`.

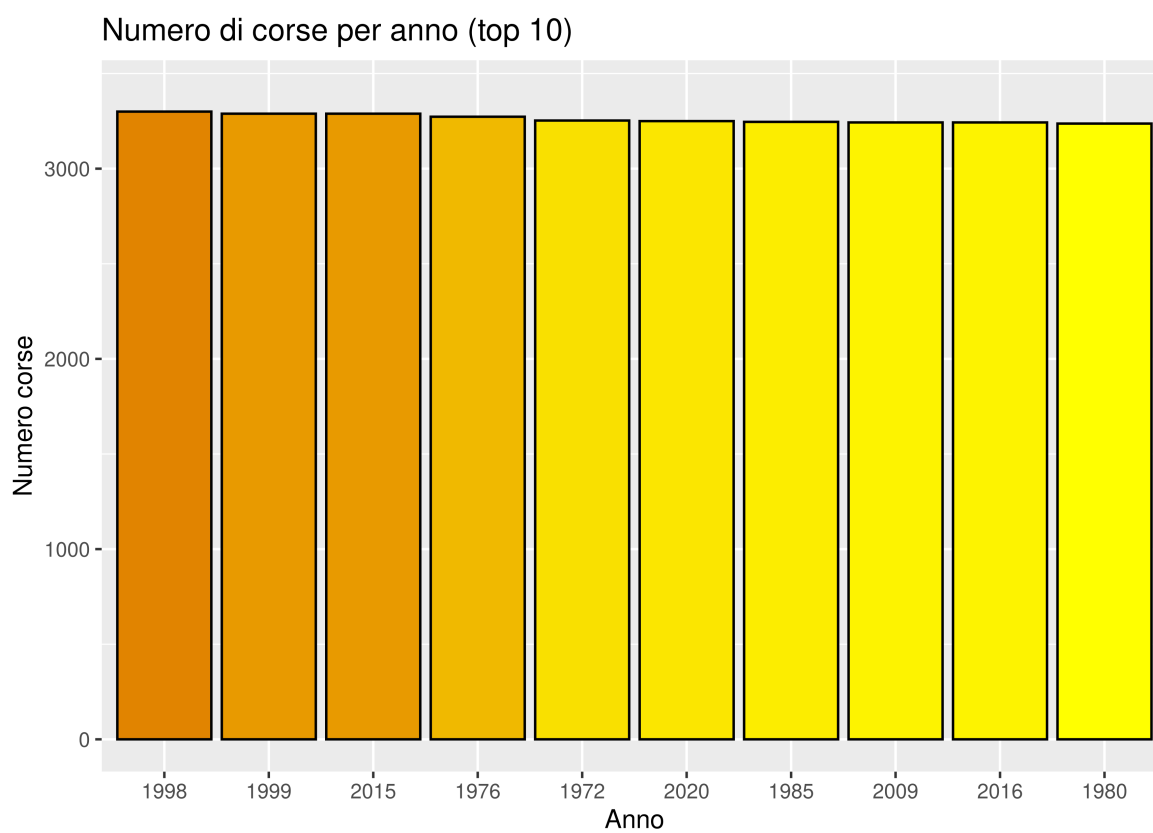
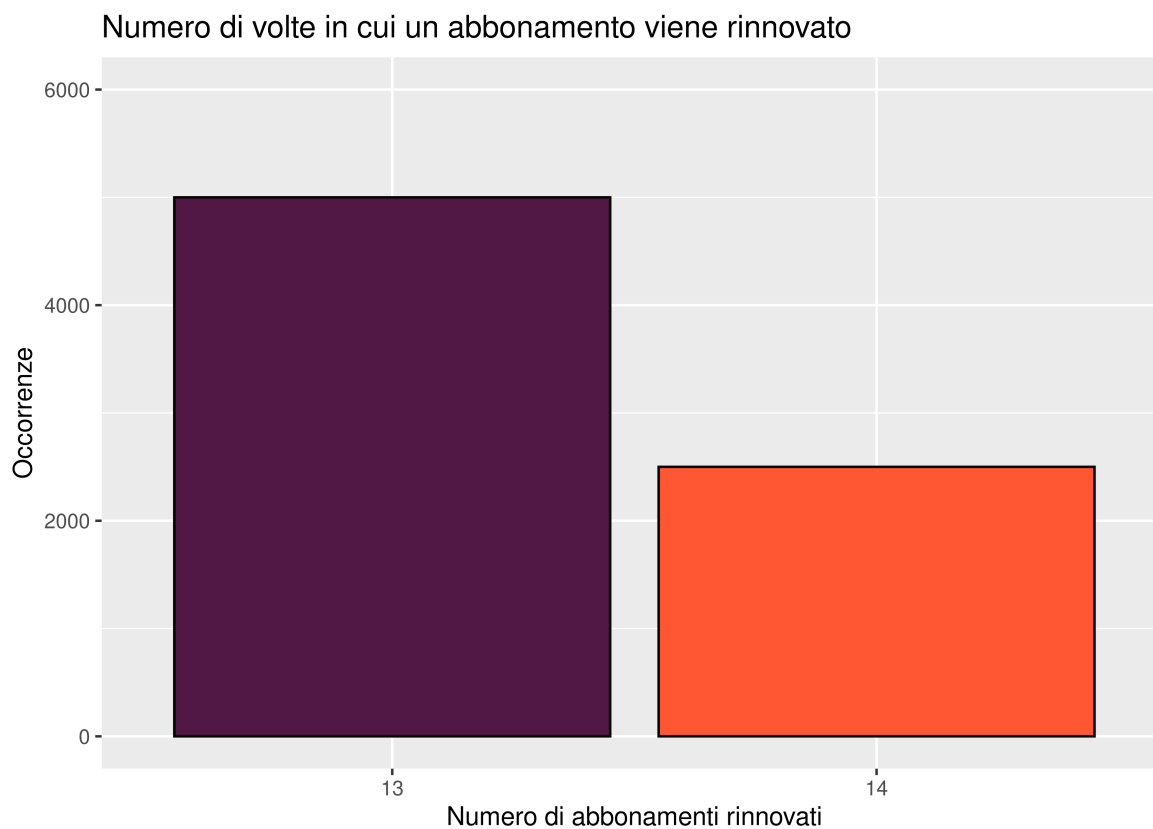
Si noti che si sono sostituiti i 3 attributi che erano chiave esterna in `Corsa` con un singolo ID, proveniente da `HaEseguito`. Ciò è stato fatto al fine di evitare di avere troppi attributi come chiavi esterne di `Corsa`.

Non è disponibile, inoltre, il codice per il popolamento del database, il quale viene generato in automatico dallo script python (che può essere trovato in `scripts/populate.py`).

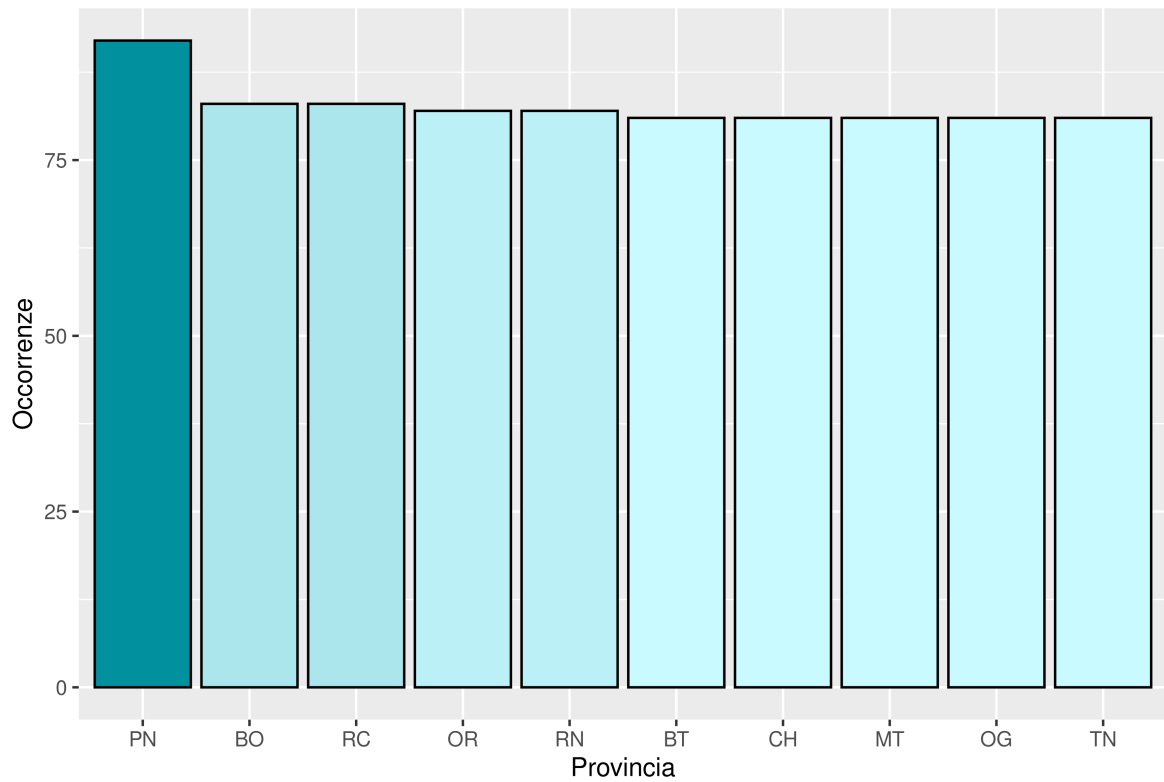
Infine, per mantenere le relazioni `valido` e `scaduto` coerenti, si è utilizzato il tool `pgcron` per effettuare gli spostamenti necessari ad intervalli di tempo regolari.

7 Analisi dei dati

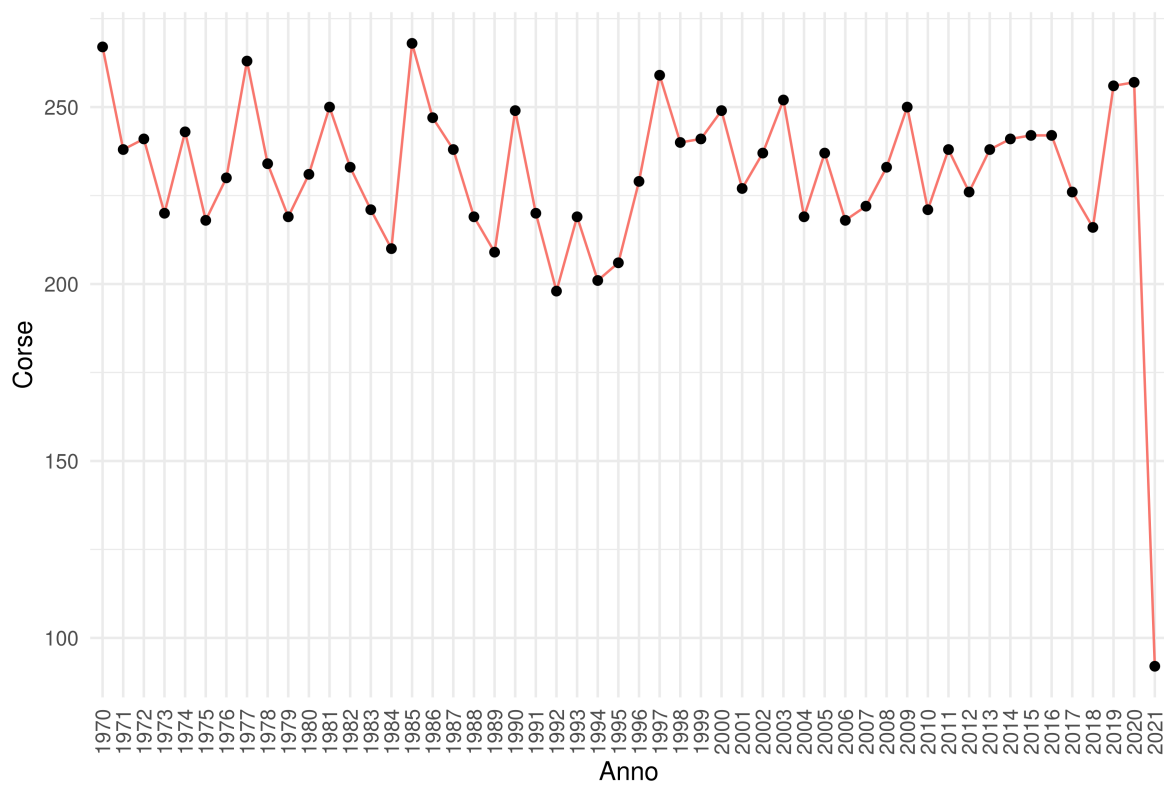
7.1 Query



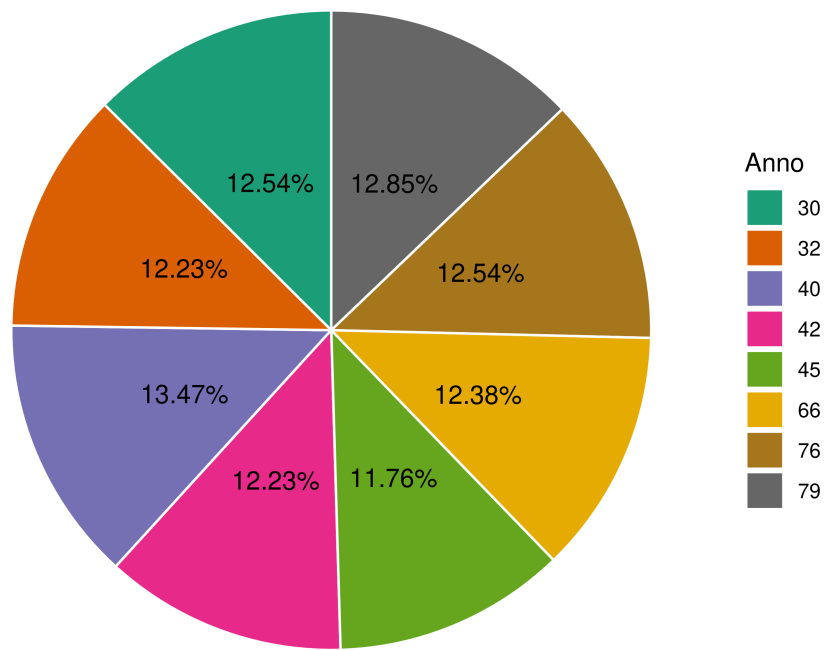
Province di appartenenza dei clienti (top 10)



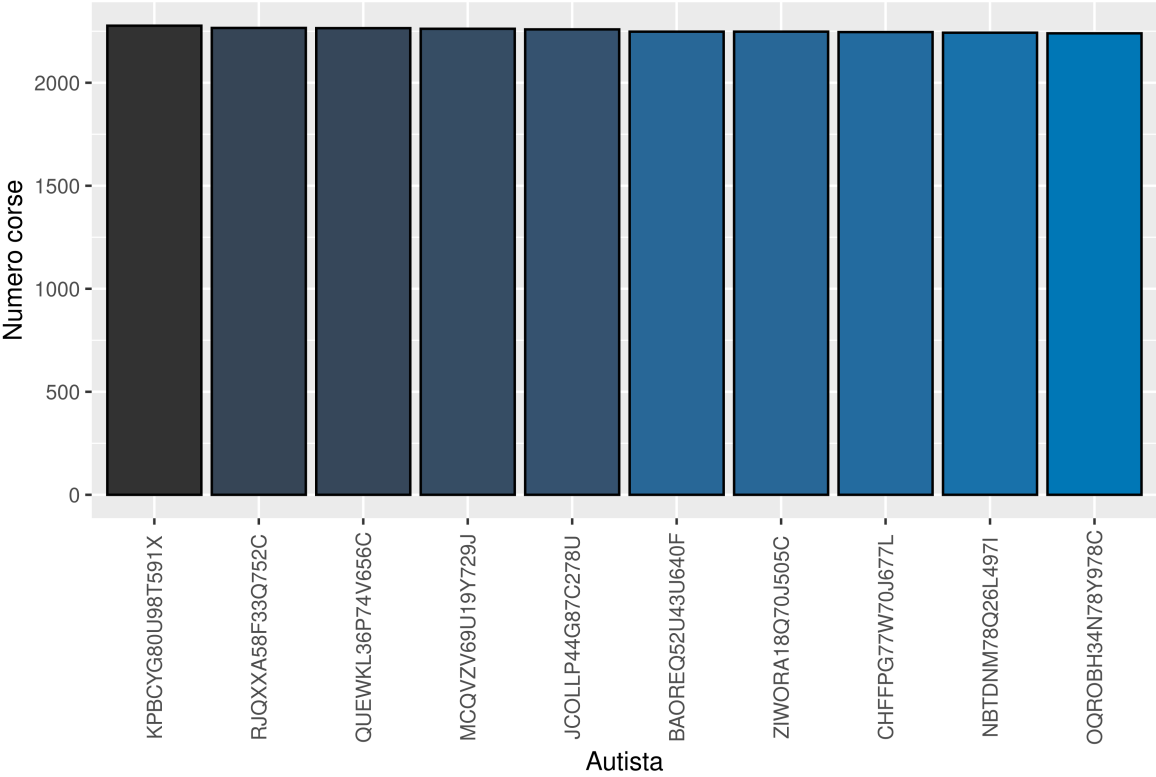
Numero di corse per ogni anno



Età dei clienti (top 8)

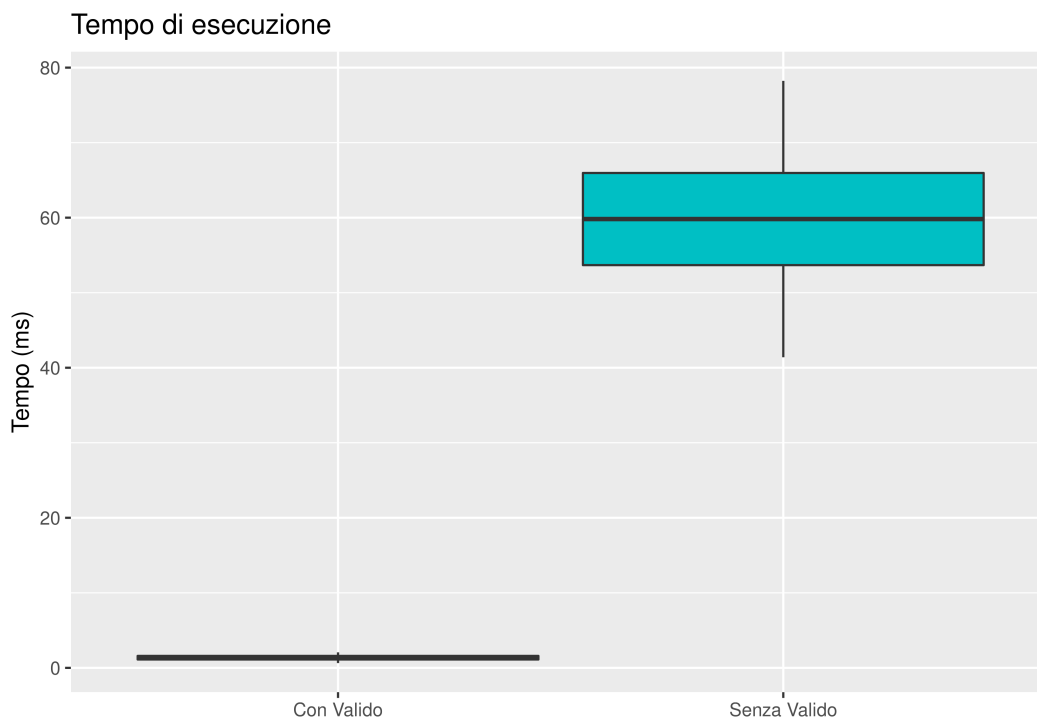
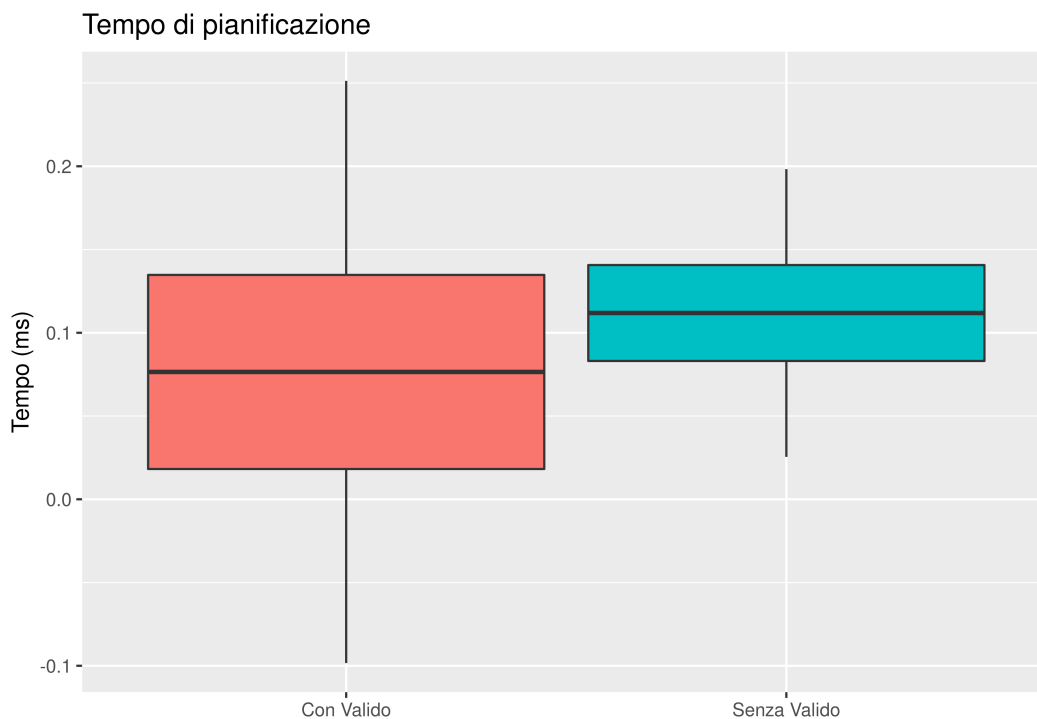


Numero di corse per autista (top 10)



7.2 Performance

Si è analizzata la performance della scelta progettuale effettuata nella [sezione 4.2.3](#). Si riportano di seguito i relativi grafici:



Si è inoltre tentato di analizzare l'efficacia dell'indice definito nella [sezione 5](#). Questo non è stato possibile in quanto **PostgreSQL non permette di forzare l'utilizzo di un indice** ma permette solamente di indicargli l'uso se lo ritiene opportuno (tramite opzioni come `SET enable_seqscan = OFF` o altre descritte a [questa](#)).