Final Project Report

Student: Fuad Ismayilbayli
Project Topic: Analysis of Estonian Startups

Introduction

For this project, I used a CSV file named estonia_startapp.csv, which contains information about Estonian startups, including:

- name – company name
- code – unique company identifier
- type – type of company (e.g., Osaühing, Mittetulundusühing)
- kmkr_nr – VAT number
- status – company status
- open_date – date of registration
- adress – street address
- city – city of registration

This CSV file was imported into a SQLite database (new_startups.db) using Python.

1 – Problem Description (30%)

Problem Background

This project uses real-world data from the CSV file estonia_startapp.csv, which contains detailed information about Estonian startups. The goal is to design a properly normalized relational database based on this dataset, load it into SQLite, and perform meaningful analytical SQL queries.

The CSV was imported into a SQLite database (new_startups.db) using Python. After importing, the data was normalized into 6 entities according to relational database design principles.

Users

- **Business analysts studying startup growth**
- **Government agencies analyzing economic activity**
- **Investors evaluating startup ecosystems**
- **Researchers collecting registration statistics**

Entities and Relationships

This database contains six entities:

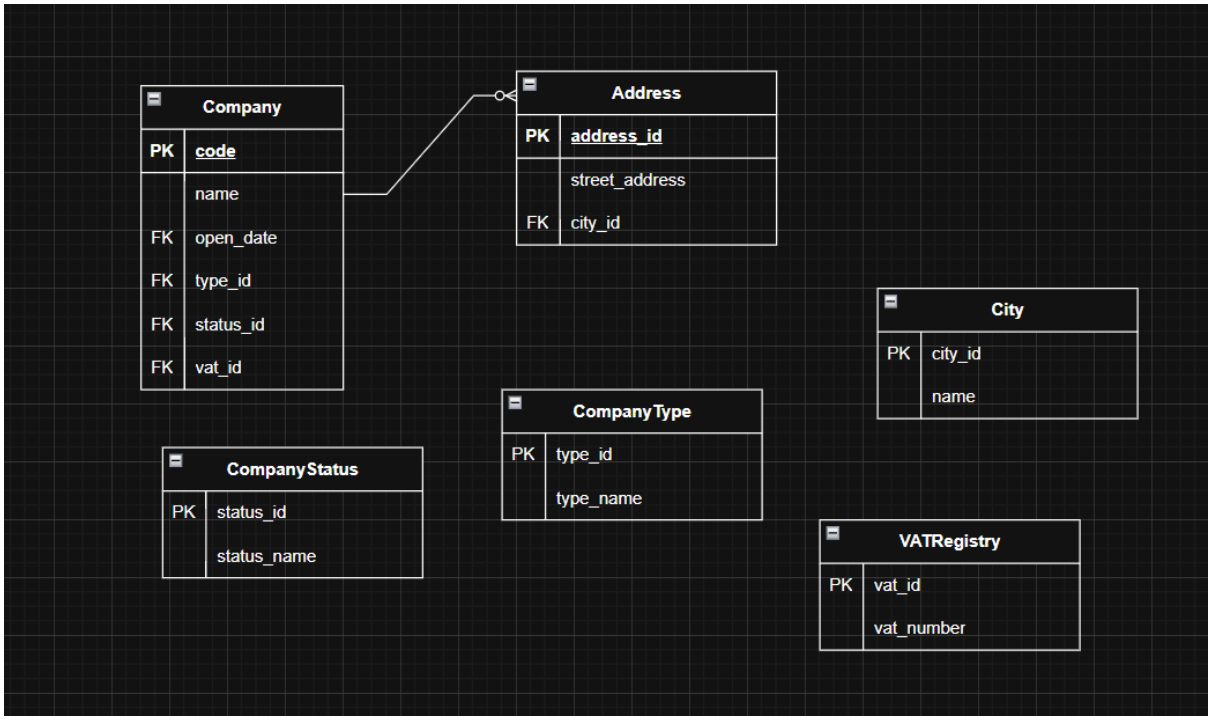
1. **Company**
2. **Address**
3. **City**
4. **CompanyType**
5. **CompanyStatus**
6. **VATRegistry**

Business Rules

1. **Each company must belong to exactly one company type.**
2. **Each company must have exactly one status.**
3. **Each company must be registered at exactly one address.**
4. **Each address must belong to exactly one city.**
5. **VAT numbers are stored separately and each company can have zero or one VAT record.**

These rules enforce consistency, reduce redundancy, and allow flexible expansion of the database.

2 – ERD Diagram (30%)



Normalization

- **All lookup data (type, status, city, VAT) moved to separate tables.**
- **Ensures 3rd Normal Form (3NF).**
- **Eliminates repeated city names, repeated company types, and repeated statuses.**

3 – Screenshots & Comments (40%)

SQL Queries and Results:

1)List of companies with addresses (JOIN)

- **SQL:** first.sql
- **Result:** first_result.txt
- **Comment:**
    - Performs an INNER JOIN between Company and Address to show complete information.
    - The first 50 companies are displayed to demonstrate table relationships.

IT-NKTU > database > first.sql

```sql
-- SQLite
SELECT c.name, c.code, c.type, c.open_date, a.street_address, a.city
FROM Company c
JOIN Address a ON c.address_id = a.address_id
ORDER BY c.open_date
LIMIT 50;
```

Result:

third_result.txt | second_result.txt | SQLite | first_result.txt | third.sql | first.sql | second.sql

SQL ▾ 1 / 1 1 - 50 of 50

name	code	type	open_date	street_address	city
Anija Valla Noortekeskus	75039770	Kohaliku omavalitsuse asutus	01.01.2017	F. R. Kreutzwaldi tn 2	Kehra linn, Anija vald, Harju maakond
Avo Naar	10644945	Füüsilisest isikust ettevõtja	01.02.2000		Penu küla, Häädemeeste vald, Pärnu maakond
Borealis Eesti OÜ	11100242	Osaühing	01.02.2005	Lõõtsa tn 5 // Sepapaja tn 4	Lasnamäe linnaosa, Tallinn, Harju maakond
Businessline OÜ	11214595	Osaühing	01.02.2006	Sõpruse pst 145	Kristiine linnaosa, Tallinn, Harju maakond
A&T Prisce OÜ	11214891	Osaühing	01.02.2006	Pikk jalg 16	Kesklinna linnaosa, Tallinn, Harju maakond
Asset Pro OÜ	11215070	Osaühing	01.02.2006	Uus tn 24-7	Kesklinna linnaosa, Tallinn, Harju maakond
Antiiv Invest OÜ	11334566	Osaühing	01.02.2007	E. Vilde tn 13-33	Rakvere linn, Lääne-Viru maakond
Arlop Grupp OÜ	11344671	Osaühing	01.02.2007	Lageda-3	Kõljala küla, Saaremaa vald, Saare maakond
Aroz3d OÜ	11875084	Osaühing	01.02.2010	Rähni tn 35	Rakvere linn, Lääne-Viru maakond
Aritmeetika OÜ	11883988	Osaühing	01.02.2010	Kaarli tee 6	Sõrve küla, Harku vald, Harju maakond
AS PANAVIATIC Maintenance	11884054	Aktsiaselts	01.02.2010	Väike-Sõjamäe tn 22	Lasnamäe linnaosa, Tallinn, Harju maakond
Boruti OÜ	11884215	Osaühing	01.02.2010	C. R. Jakobsoni tn 21-5	Kesklinna linnaosa, Tallinn, Harju maakond
Autoremont Alo Sohar	11889169	Füüsilisest isikust ettevõtja	01.02.2010	Jaama tn 49	Põlva linn, Põlva vald, Põlva maakond
Bryoria OÜ	12052462	Osaühing	01.02.2011	Uuemäe	Puusepa küla, Kose vald, Harju maakond
Audio System Design OÜ	12052491	Osaühing	01.02.2011	Västriku tn 8-500	Kristiine linnaosa, Tallinn, Harju maakond
Befle OÜ	12052671	Osaühing	01.02.2011	Lennuki tn 4-2	Rakvere linn, Lääne-Viru maakond
Baltic Floor OÜ	12228725	Osaühing	01.02.2012	Riia mnt 51-4	Pärnu linn, Pärnu linn, Pärnu maakond
BCC RAAMATUPIDAMINE OÜ	12415261	Osaühing	01.02.2013	Roseni tn 13	Kesklinna linnaosa, Tallinn, Harju maakond
ARGO&COMPANI OÜ	12415502	Osaühing	01.02.2013	Sadama tn 32	Paldiski linn, Lääne-Harju vald, Harju maakond
Artur Stefani	12987230	Füüsilisest isikust ettevõtja	01.02.2016	Olevi tn 23-28	Järve linnaosa, Kohtla-Järve linn, Ida-Viru maakond
Art Kaminad OÜ	12987419	Osaühing	01.02.2016	Pärnu mnt 106-87	Kesklinna linnaosa, Tallinn, Harju maakond
A.R.C Trading OÜ	12987477	Osaühing	01.02.2016	Asula tn 1-19	Kesklinna linnaosa, Tallinn, Harju maakond

2)Number of companies by year of registration (GROUP BY)

- **SQL:** second.sql
- **Result:** second_result.txt
- **Comment:**
    - Aggregates companies by year (COUNT(*)) to show registration activity.
    - Helps identify the most active years for new startups.

IT-NKTU > database > second.sql

```sql
-- SQLite
SELECT SUBSTR(c.open_date, 7, 4) AS year, COUNT(*) AS total_companies
FROM Company c
GROUP BY year
ORDER BY total_companies DESC;
```

Result:

third_result.txt | SQLite | second_result.txt | first_result.txt | third.sql | first.sql | second.sql

SQL ▾ 1 / 1 1 - 31 of 31

year	total_companies
2024	1782
2025	1655
2023	1512
2021	1450
2019	1407
2018	1370
2022	1300
2020	1252
2017	1182
2016	1063
2015	845
2014	759
2009	755
2013	704
2012	697
2011	607
2010	542
2007	339
2008	310
2006	283
1997	276

3)Top 3 cities with most companies (window function)

- **SQL:** third.sql
- **Result:** third_result.txt
- **Comment:**

    - Finds the top 3 cities with the highest number of companies.
    - Uses a window function (RANK() OVER) to rank companies within each city by registration date.
    - Demonstrates advanced SQL features such as CTEs and window functions.

IT-NKTU > database > third.sql

```sql
WITH city_counts AS (
    SELECT a.city, COUNT(*) AS total_companies
    FROM Company c
    JOIN Address a ON c.address_id = a.address_id
    GROUP BY a.city
),
top_cities AS (
    SELECT city
    FROM city_counts
    ORDER BY total_companies DESC
    LIMIT 3
),
ranked AS (
    SELECT c.name, c.open_date, a.city,
           RANK() OVER (PARTITION BY a.city ORDER BY c.open_date) AS rank_in_city
    FROM Company c
    JOIN Address a ON c.address_id = a.address_id
    WHERE a.city IN (SELECT city FROM top_cities)
)
SELECT *
FROM ranked
WHERE rank_in_city <= 5
ORDER BY city DESC, rank_in_city;

```

Result:

SQL ▾ 1 / 1 1 - 18 of 18

name	open_date	city	rank_in_city
Baudit KVA OÜ	01.03.2024	Tartu linn, Tartu linn, Tartu maakond	1
Anneli Loodus	01.04.2010	Tartu linn, Tartu linn, Tartu maakond	2
Blehner OÜ	01.04.2013	Tartu linn, Tartu linn, Tartu maakond	3
Auland OÜ	01.04.2020	Tartu linn, Tartu linn, Tartu maakond	4
Barber Taali OÜ	01.04.2024	Tartu linn, Tartu linn, Tartu maakond	5
Borealis Eesti OÜ	01.02.2005	Lasnamäe linnaosa, Tallinn, Harju maakond	1
AS PANAVIATIC Maintenance	01.02.2010	Lasnamäe linnaosa, Tallinn, Harju maakond	2
ASKE OÜ	01.02.2016	Lasnamäe linnaosa, Tallinn, Harju maakond	3
Beluga4Health OÜ	01.02.2019	Lasnamäe linnaosa, Tallinn, Harju maakond	4
Arsa OÜ	01.02.2019	Lasnamäe linnaosa, Tallinn, Harju maakond	4
A&T Prisce OÜ	01.02.2006	Kesklinna linnaosa, Tallinn, Harju maakond	1
Asset Pro OÜ	01.02.2006	Kesklinna linnaosa, Tallinn, Harju maakond	1
Boruti OÜ	01.02.2010	Kesklinna linnaosa, Tallinn, Harju maakond	3
BCC RAAMATUPIDAMINE OÜ	01.02.2013	Kesklinna linnaosa, Tallinn, Harju maakond	4
Art Kaminad OÜ	01.02.2016	Kesklinna linnaosa, Tallinn, Harju maakond	5
A.R.C Trading OÜ	01.02.2016	Kesklinna linnaosa, Tallinn, Harju maakond	5
BV INVEST OÜ	01.02.2016	Kesklinna linnaosa, Tallinn, Harju maakond	5
Areta OÜ	01.02.2016	Kesklinna linnaosa, Tallinn, Harju maakond	5

ZIP File Contents

File Descriptions

Database
Folder containing the SQLite database, CSV source file, and all SQL scripts/results.
estonia_startapp.csv
The original dataset containing Estonian startup information.
Used as the source file for importing data into the SQLite database.
new_startups.db
The SQLite database created for the project.
Contains the normalized tables and all imported startup data.
first.sql
SQL script for Query #1 (JOIN query).
Demonstrates retrieving company information together with related address and city details.
first_result.txt
Text file containing the output of Query #1 as executed on the database.
second.sql
SQL script for Query #2 (aggregation / GROUP BY).
Counts the number of companies grouped by their registration year.
second_result.txt
Text file containing the output of Query #2.
third.sql
SQL script for Query #3 (window function).
Shows the top cities by number of companies using ranking (ROW_NUMBER / DENSE_RANK).
third_result.txt
Text file containing the output of Query #3.

Summary:

- Database was created from the CSV file, normalized into two tables: Company and Address.
- Three queries demonstrate: JOIN, aggregation (GROUP BY), and window functions (RANK).
- Query results and scripts are included in the ZIP file for submission.