

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150
ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm



ETL ASSIGNMENT PRESENTATION

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

PURPOSE STATEMENT

This document provides a presentation overview of an ETL assignment issued by CROZ. It is intended solely for assessment of candidate's technical knowledge and expertise.

DISCLAIMER

This document in any form, software or printed matter, contains proprietary information and sensitive data that is the exclusive property of the candidate; Zvonimir Klenovic. This document and information contained herein may not be disclosed, copied, reproduced or distributed to anyone outside CROZ without prior written consent of Zvonimir Klenovic.

This document is for informational purposes only and is intended solely to assist you in the selection process of the candidate.



PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

ZVONIMIR KLENOVIC

master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

TABLE OF CONTENTS

Purpose Statement	1
Disclaimer	1
Table of Contents	2
Introduction	3
Assignment	3
Technologies used	3
Sources	3
Talend	4
ETL Process Sketch	5
E/R Model Sketch	6
Extract	7
Transform	8
Load	10
SQL Queries	12
Analytics	14
Jupyter Notebook	16
Jupyter Notebook Extra	19
Conclusion	20

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

INTRODUCTION

This presentation will demonstrate my abilities to handle tasks such as the ones issued in CROZ assignment. I will focus on providing practical solutions to various types of problems which can be encountered in the ETL process; either system, software or task itself related.

The presented problems will be addressed with images, code syntax and or additional external links or files.

ASSIGNMENT

The assignment asks for a creation of a simple ETL process followed by creating Analytical reports:

- 2 different sources of Database
- perform JOIN in Talend Open Source Tool
- output and store new joined database in Oracle XE or Oracle DWH
- perform additional analytical SQL queries
- present Analytic reports accompanied with Visualizations

TECHNOLOGIES USED

MS Excel 2010
MySQL 8.0.17
JDK-8u221

Talend 7.2.1
Oracle Database 19c
SQL Developer 19.2.1

Jupyter Notebook

SOURCES

- Nordea Bank statement for the period 01.01.-28.09.2019
downloaded as Excel and imported into MySQL

Open a MySQL Connection-Export Excel Data to New Table

Possible problems occurring with columns not displaying correctly can be fixed by either selecting **First Row Contains Column Names** or a problem with .csv files changing delimiter settings in either Excel itself or Windows system's regional settings, it may expect a semicolon as a separator instead of a comma, since in some languages the comma is used as the decimal separator.

- Daily currency exchange rate from HNB
for the period equal to bank statement
(01.01.-28.09.2019)

<https://www.hnb.hr/temeljne-funkcije/monetarna-politika/tecajna-lista/tecajna-lista>

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

TALEND v7.2.1

Installation

Keep in mind that different versions of Talend work with different versions of JDK. Possible problems with starting Talend include not being able to start Virtual Machine. This is a Java related error which can be fixed by editing [Environment Variables in Windows Advanced System Settings](#).

Edit PATH-Add-C:\Program Files\Java\jdk1.8.0_221\bin

Now we can check in cmd

```
C:\Users\pizz>javac -version
javac 1.8.0_221
C:\Users\pizz>
```

Some other system related Tweaks which seem to work for older versions of Talend include editing the .INI file located in Talend install dir.

Creating a MySQL Connection for first SOURCE Database resulted with a problem.

java.sql.SQLException:
The server time zone value 'CEST' is unrecognized or represents more than one time zone.
You must configure either the server or JDBC driver (via the serverTimezone configuration property) to use a more specific time zone value if you want to utilize time zone support

Issue can be solved by adding

```
useUnicode=true&useJDBCCompliantTimezoneShift=true&useLegacyDatetimeCode=false&
serverTimezone=UTC
```

in the [Additional parameters](#) section while creating a Database connection on repository.

Second SOURCE database was added as input file Delimited with Display First Row as columns.

Creating an Oracle connection resulted in downloading and installing a specific .jar driver requested by Talend.

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

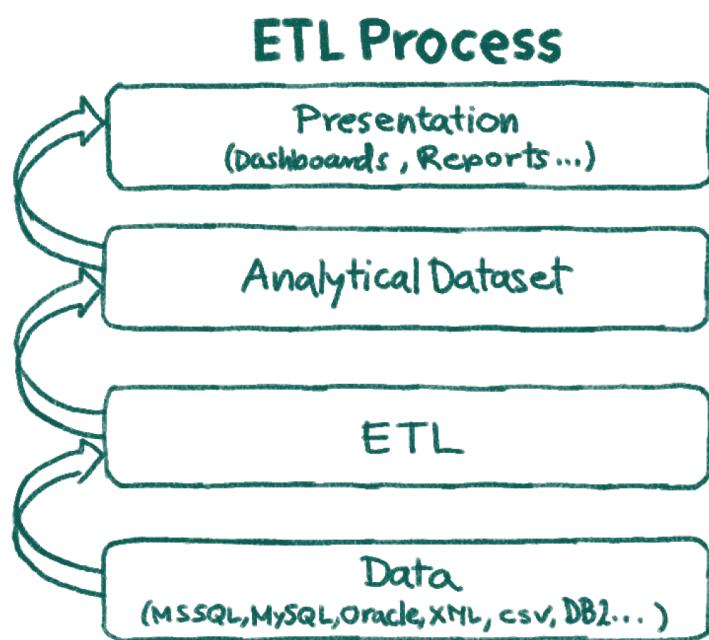
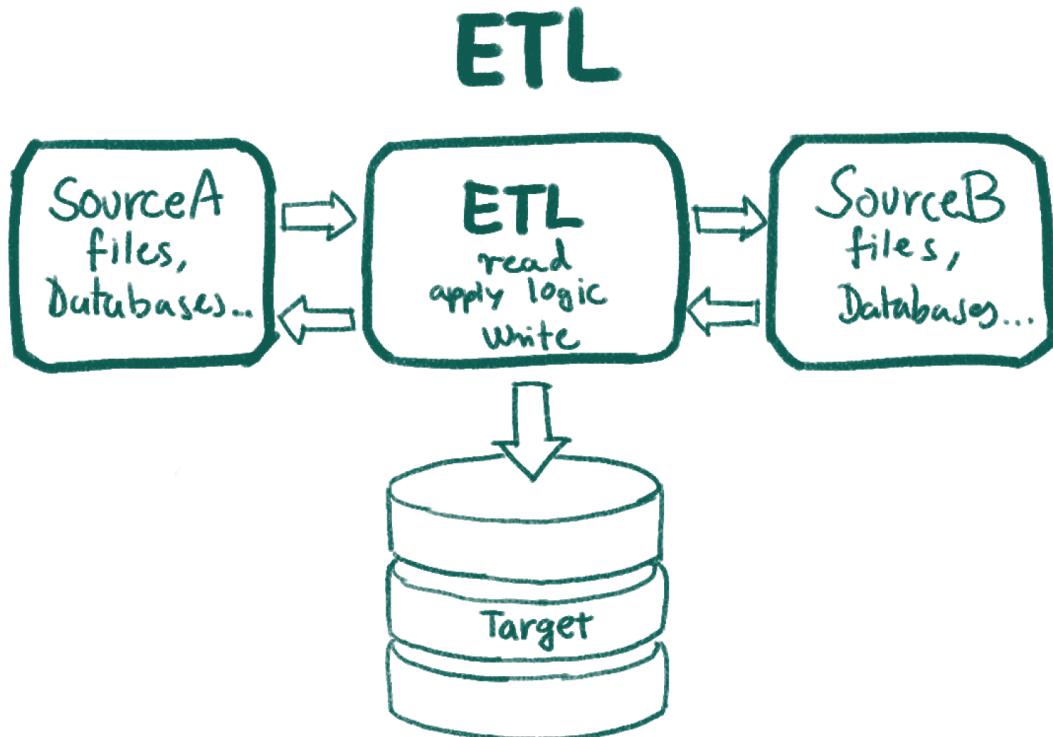
**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

ETL PROCESS SKETCH

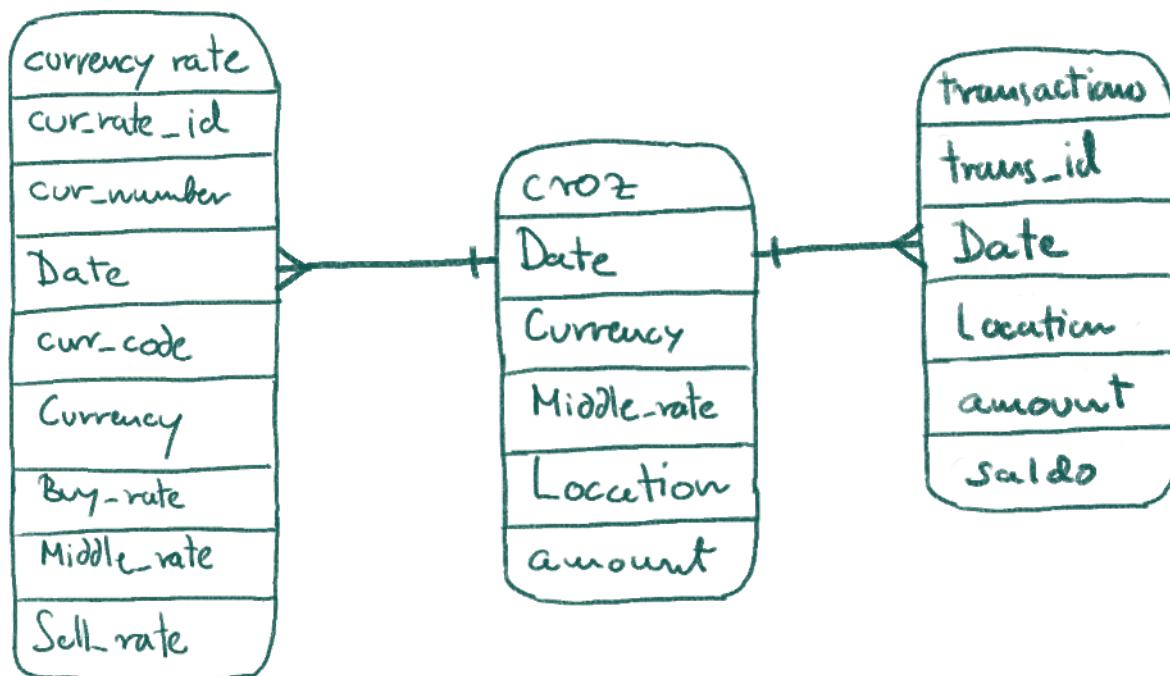
I took the liberty of drawing a freehand sketch done in Corel Painter showing my vision of ETL Process .



E/R MODEL SKETCH

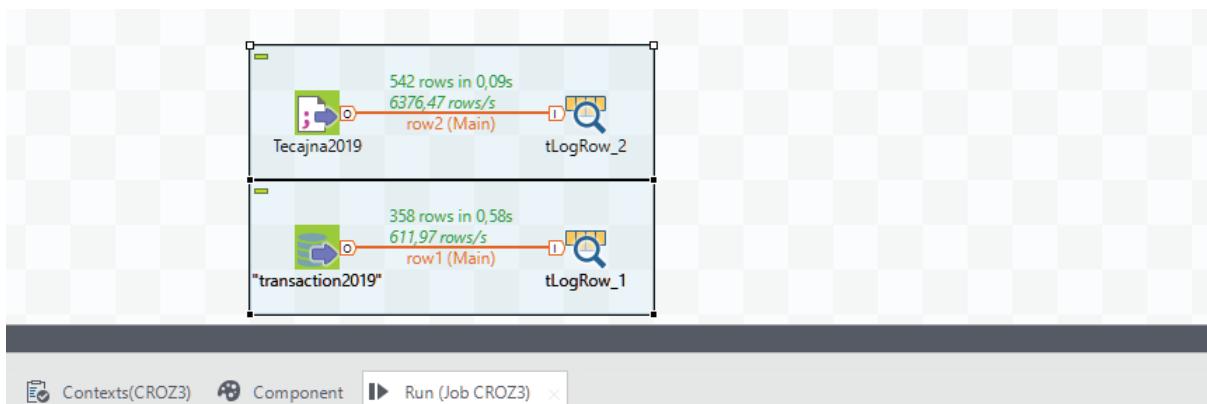
Once again I took the liberty of drawing a freehand sketch. I am not completely sure if this is the exact sketch you were looking for but I can say is that the reason I have choosed One-To-Many connections is because "Date" can have a multitude of transactions and currencies attached.

I also believe that "transactions" are a weak entity because they would be dependant on client.



EXTRACTION

With the two sources connected I was ready to extract the data. I had created a new JOB in Talend, selected MySQL connection and inputFileDelimited as my input components and connected each one to a tLogRow component to oversee successful extraction.



Execution

Run Kill Clear

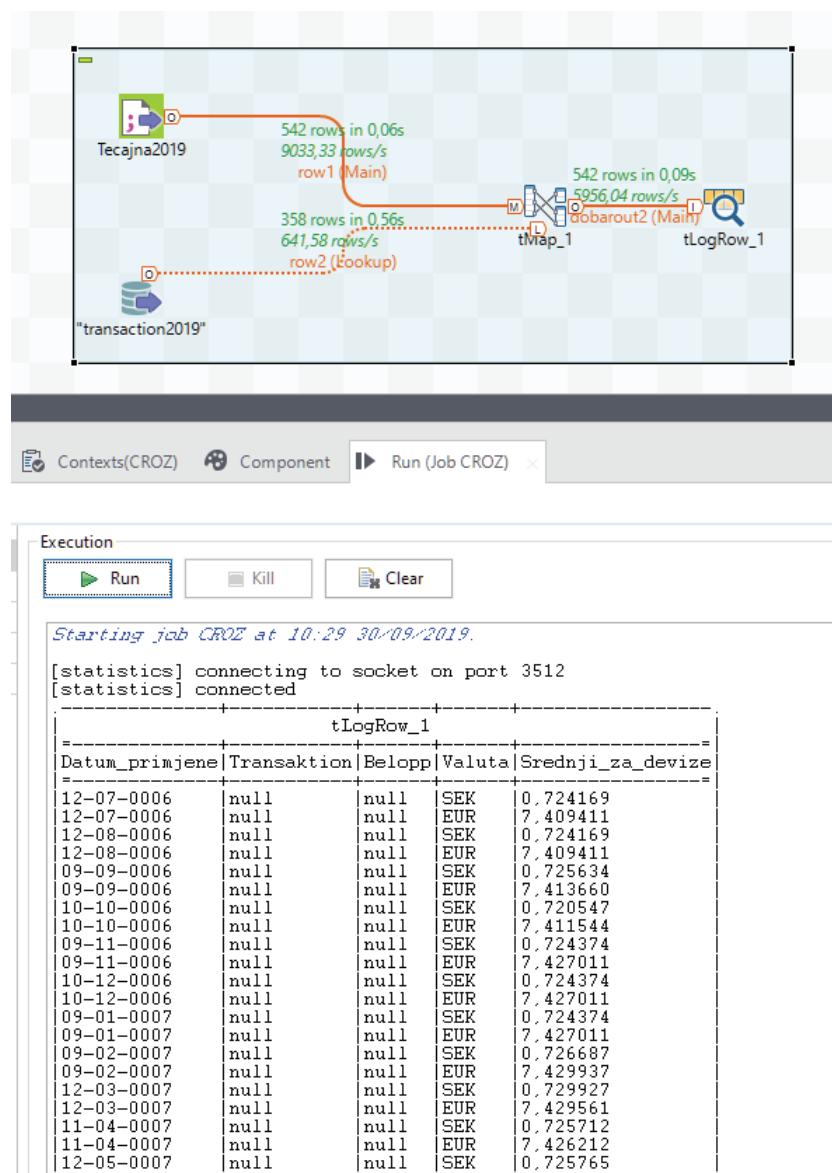
14-01-2019 Kortköp 190112 ROT BUTIK O KÖK -221,30	14-01-2019 Kortköp 190112 HEMKÖP STOCKHOLM GUL -213,20
14-01-2019 Kortköp 190111 DEN GRONE JAGAREN -116,00	14-01-2019 Kortköp 190111 DEN GRONE JAGAREN -174,00
11-01-2019 Kortköp 190109 ICA SUPERMARKET GLOB -135,50	11-01-2019 Kortköp 190109 HEMKÖP STOCKHOLM GUL -264,65
11-01-2019 Kortköp 190109 Babbel.com -99,00	11-01-2019 Kortköp 190108 HEMKÖP STOCKHOLM GUL -358,55
10-01-2019 Kortköp 190108 HEMKÖP STOCKHOLM GUL 5.000,00	08-01-2019 MARTINA.KUCA -178,75
02-01-2019 Kortköp 181227 HEMKÖP STOCKHOLM GUL -87,60	02-01-2019 Kortköp 181228 ICA SUPERMARKET GLOB -12,00
01-01-2019 Vardagspaket Tillval Betalningar -24,00	01-01-2019 Vardagspaket Månadspris kort -24,00

tLogRow_2

Column0	Broj_te_ajnice	Datum_primjene	Dr_ava	Ifra_valute	Valuta	Jedinica	Kupovni_za_devize	Srednji_z
0	1	12-07-0006	Ivedska	752	SEK	1	0,721996	0,724169
1	1	12-07-0006	EMU	978	EUR	1	7,387183	7,409411
2	1	12-08-0006	Ivedska	752	SEK	1	0,721996	0,724169
3	1	12-08-0006	EMU	978	EUR	1	7,387183	7,409411
4	2	09-09-0006	Ivedska	752	SEK	1	0,723457	0,725634
5	2	09-09-0006	EMU	978	EUR	1	7,391419	7,413660
6	3	10-10-0006	Ivedska	752	SEK	1	0,718385	0,720547
7	3	10-10-0006	EMU	978	EUR	1	7,389309	7,411544
8	4	09-11-0006	Ivedska	752	SEK	1	0,722201	0,724374
9	4	09-11-0006	EMU	978	EUR	1	7,404730	7,427011
10	4	10-12-0006	Ivedska	752	SEK	1	0,722201	0,724374
11	4	10-12-0006	EMU	978	EUR	1	7,404730	7,427011
12	4	09-01-0007	Ivedska	752	SEK	1	0,722201	0,724374
13	4	09-01-0007	EMU	978	EUR	1	7,404730	7,427011

TRANSFORMATION

Here I had encountered a problem off all values from one table being displayed as NULL after join.



Only matching columns in both tables were the Date columns (Datum and Datum_primjene) so it had seemed like a logical selection to preform a join. The problem was that Date columns were in different formats.

TRANSFORMATION

Manually I tried configuring source input schemas in Date columns to match by type and pattern.
I started a topic in Talend Community describing my problem.

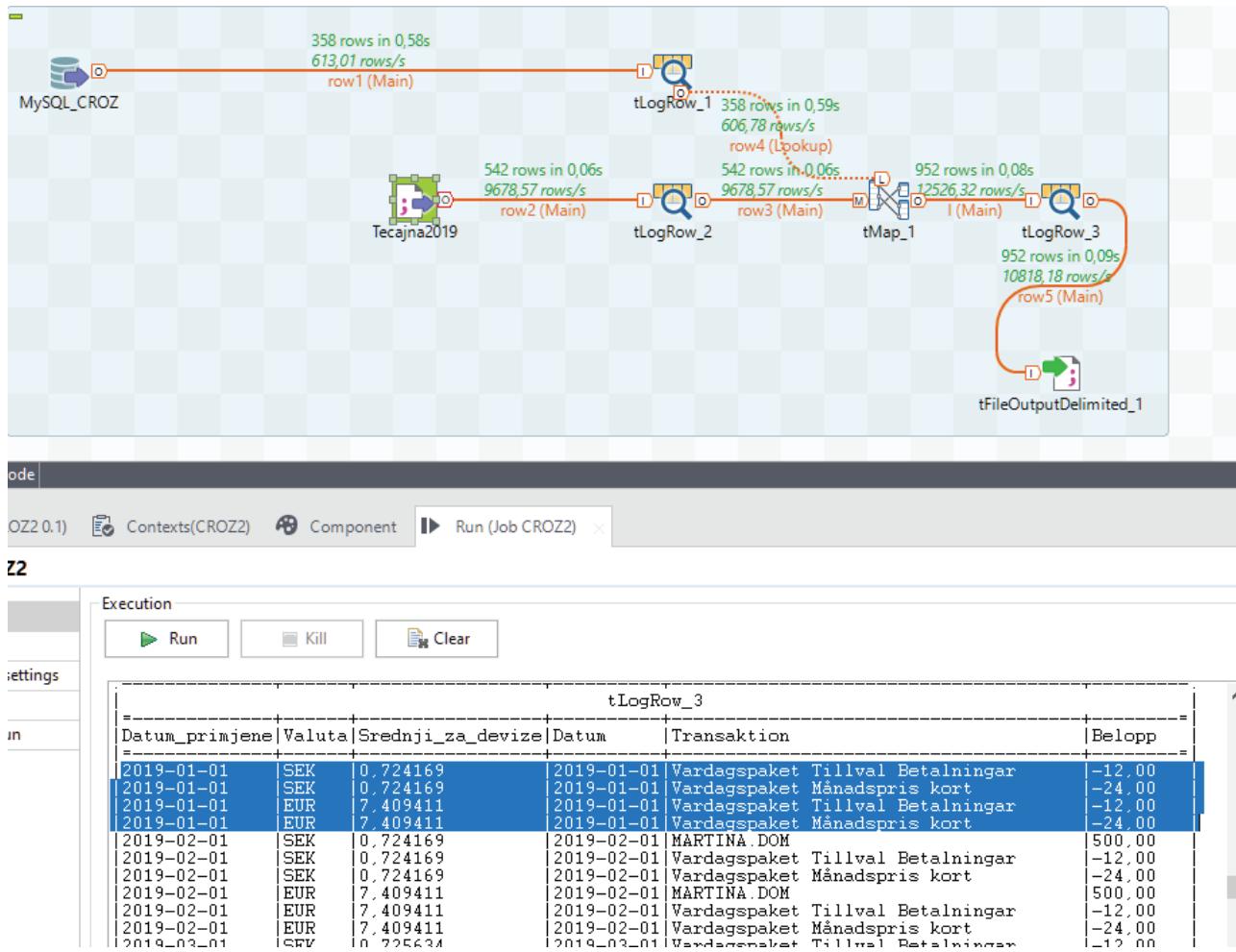
<https://community.talend.com/t5/Design-and-Development/tMap-Problem-with-NULL-values-on-identic-JOBS/m-p/203356>

Even tough hints were suggesting I should convert both columns to DATE without any pattern that did not result in a successful join. I had even tried to convert columns before importing into Talend.

e.g. I used SQL syntax for transactions database in MySQL.

```
ALTER TABLE mydb.transactions  
MODIFY COLUMN mydb.transactions.date DATE;
```

Also unssucessful. Eventually I came up with the idea to match date patterns in built in schemas for Date columns of both sources and then convert them to STRING. I was now able to do a successful JOIN. In the tMap component I had chosen to MATCH ALL ROWS in an OUTER JOIN od Date columns in both databases and output a new database as a Delimited File with all necessary columns.

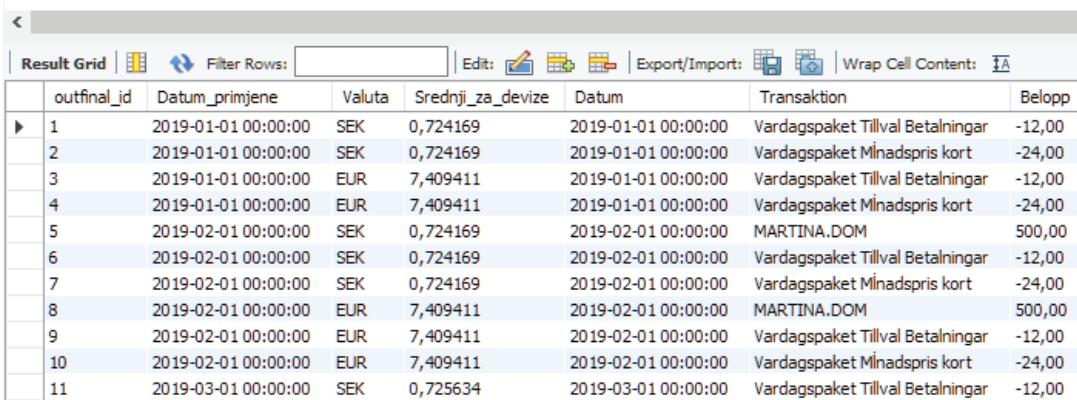


LOAD

For practice I first loaded the resulting DataBase again in MySQL and performed a few queries.

e.g. Displaying the new DataBase

```
1 •  SELECT * FROM sakila.outfinal;  
2  
3  
4
```

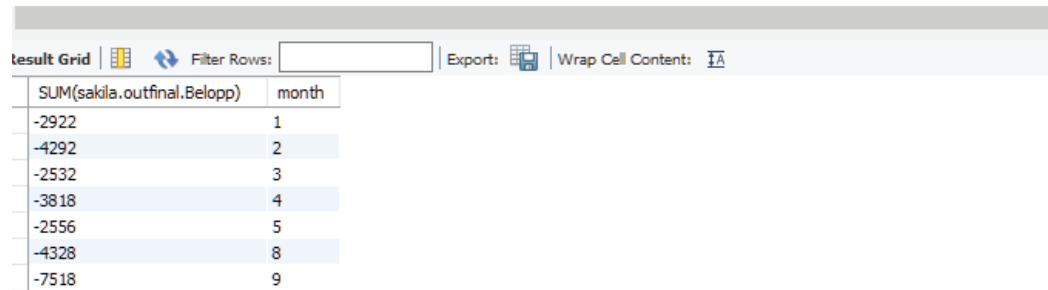


	outfinal_id	Datum_primjene	Valuta	Srednji_za_devize	Datum	Transaktion	Belopp
▶	1	2019-01-01 00:00:00	SEK	0,724169	2019-01-01 00:00:00	Vardagspaket Tillval Betalningar	-12,00
	2	2019-01-01 00:00:00	SEK	0,724169	2019-01-01 00:00:00	Vardagspaket Mlnadspris kort	-24,00
	3	2019-01-01 00:00:00	EUR	7,409411	2019-01-01 00:00:00	Vardagspaket Tillval Betalningar	-12,00
	4	2019-01-01 00:00:00	EUR	7,409411	2019-01-01 00:00:00	Vardagspaket Mlnadspris kort	-24,00
	5	2019-02-01 00:00:00	SEK	0,724169	2019-02-01 00:00:00	MARTINA.DOM	500,00
	6	2019-02-01 00:00:00	SEK	0,724169	2019-02-01 00:00:00	Vardagspaket Tillval Betalningar	-12,00
	7	2019-02-01 00:00:00	SEK	0,724169	2019-02-01 00:00:00	Vardagspaket Mlnadspris kort	-24,00
	8	2019-02-01 00:00:00	EUR	7,409411	2019-02-01 00:00:00	MARTINA.DOM	500,00
	9	2019-02-01 00:00:00	EUR	7,409411	2019-02-01 00:00:00	Vardagspaket Tillval Betalningar	-12,00
	10	2019-02-01 00:00:00	EUR	7,409411	2019-02-01 00:00:00	Vardagspaket Mlnadspris kort	-24,00
	11	2019-03-01 00:00:00	SEK	0,725634	2019-03-01 00:00:00	Vardagspaket Tillval Betalningar	-12,00

Trying out more complex queries

e.g. Displaying a list of total spent in a specific store per month etc.

```
1 •  SELECT SUM(sakila.outfinal.Belopp), extract(month FROM sakila.outfinal.Datum_primjene)  
2   AS month  
3   FROM sakila.outfinal  
4   WHERE sakila.outfinal.Transaktion LIKE ('%HEM%')  
5   GROUP BY month  
6   ORDER BY month;  
7  
8  
9  
10  
11  
12  
13
```



	SUM(sakila.outfinal.Belopp)	month
	-2922	1
	-4292	2
	-2532	3
	-3818	4
	-2556	5
	-4328	8
	-7518	9

LOAD

I had encountered serious SYSTEM problems while trying to install Oracle XE 18c. The installation resulted in a BSOD which even messed up my boot order in BIOS. To completely remove this version I had to manually delete all Windows REGISTRY entries for Oracle together with Admin and Services. Installation of Oracle Database 19c went smoothly. A successful connection was established with Talend and now I have installed Oracle SQL Developer 19.2.1 to create my schema and load into Oracle DB.

Since my Output database columns have been converted to STRING before I had to create a matching schema in SQL Developer.

The screenshot shows the Oracle SQL Developer interface. On the left, there's a tree view of tables under 'Tables (Filtered)'. Some tables listed are AQ\$_INTERNET_AGENT_PRIVS, AQ\$_INTERNET_AGENTS, AQ\$_KEY_SHARD_MAP, AQ\$_QUEUE_TABLES, AQ\$_QUEUES, AQ\$_SCHEDULES, CROZ_FINAL, CROZ2, and HELP. On the right, a table definition is shown with columns: COLUMN_NAME, DATA_TYPE, NULLABLE, DATA_DEFAULT, COLUMN_ID, and COMMENTS. The columns listed are: 1 DATUM_PRIMJENE, 2 VALUTA, 3 SREDNJI_ZA_DEVIZE, 4 TRANSAKTION, and 5 BELOPP.

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
1 DATUM_PRIMJENE	VARCHAR2(45 BYTE)	Yes	(null)	1	(null)
2 VALUTA	VARCHAR2(5 BYTE)	Yes	(null)	2	(null)
3 SREDNJI_ZA_DEVIZE	VARCHAR2(20 BYTE)	Yes	(null)	3	(null)
4 TRANSAKTION	VARCHAR2(45 BYTE)	Yes	(null)	4	(null)
5 BELOPP	VARCHAR2(20 BYTE)	Yes	(null)	5	(null)

Now I was certain I could load from Talend, and the JOB went through without any errors.

The screenshot shows a Talend job named 'CROZ2' in the 'Code' editor. The job flow consists of several components: MySQL_CROZ (Input), tLogRow_1, tLogRow_2, tecajna2019 (Input), tLogRow_2, tMap_1, tLogRow_3, and tDBOutput_1 (Output). The execution results window shows a table of data with columns: Date, Currency, ID, and Description. The data includes various transactions such as 'Reservation Kortköp HEMKOP STOCKHOLM G| -156,83' and 'Reservation Kortköp HEMKOP STOCKHOLM G| -14,89'. The execution statistics at the bottom show '[statistics] disconnected' and 'Job CROZ2 ended at 13:12 03/10/2019. [exit code=0]'.

Date	Currency	ID	Description
2019-09-26	EUR	7_401888	Reservation Kortköp HEMKOP STOCKHOLM G -156,83
2019-09-26	EUR	7_401888	Reservation Kortköp HEMKOP STOCKHOLM G -14,89
2019-09-26	EUR	7_401888	Reservation Kortköp ICA SUPERMARKET ME -66,57
2019-09-26	EUR	7_401888	MARTINA DOM 600,00
2019-09-26	EUR	7_401888	Kortköp 190924 7_ELEVEN 4516110 -31,00
2019-09-26	EUR	7_401888	Kortköp 190924 BARAS BACKE -29,00
2019-09-26	EUR	7_401888	Kortköp 190924 BARAS BACKE -29,00
2019-09-26	EUR	7_401888	Kortköp 190924 BARAS BACKE -29,00
2019-09-26	EUR	7_401888	Kortköp 190924 BARAS BACKE -58,00
2019-09-26	EUR	7_401888	Kortköp 190924 HEMKÖP STOCKHOLM GUI -140,59
2019-09-26	EUR	7_401888	Kortköp 190924 SHML KIOSK -59,00
2019-09-27	SEK	0_694617	Reservation Kortköp Baras Backe -58,00
2019-09-27	SEK	0_694617	Reservation Kortköp HEMKOP STOCKHOLM G -67,79
2019-09-27	SEK	0_694617	Kortköp 190925 HEMKOP STOCKHOLM GUI -80,09
2019-09-27	SEK	0_694617	Kortköp 190925 HEMKOP STOCKHOLM GUI -114,09
2019-09-27	EUR	7_405313	Reservation Kortköp Baras Backe -58,00
2019-09-27	EUR	7_405313	Reservation Kortköp HEMKOP STOCKHOLM G -67,79
2019-09-27	EUR	7_405313	Kortköp 190925 HEMKOP STOCKHOLM GUI -80,09
2019-09-27	EUR	7_405313	Kortköp 190925 HEMKOP STOCKHOLM GUI -114,09
2019-09-28	SEK	0_691671	Reservation Kortköp HONEY HONEY -54,00
2019-09-28	SEK	0_691671	Reservation Kortköp Direkten Tareqs To -59,00
2019-09-28	SEK	0_691671	Reservation Kortköp HEMKOP STOCKHOLM G -158,12
2019-09-28	EUR	7_402264	Reservation Kortköp HONEY HONEY -54,00
2019-09-28	EUR	7_402264	Reservation Kortköp Direkten Tareqs To -59,00
2019-09-28	EUR	7_402264	Reservation Kortköp HEMKOP STOCKHOLM G -158,12

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

SQL Queries

Now it was time for writing some longer SQL queries for Analytical purposes. Considering the CROZ assignment the task seemed rather simple.

I was to generate two reports:

1. A monthly statement:

- a query returning all transactions for the chosen period (I chose September)
- another query calculating SUM as TOTAL of all transactions
- and in Excel I have converted SEK to HRK by their daily exchange rates

2. Discovering clients habits:

-I decided to group transactions into 3 major groups:

SHOP

(includes queries for total monthly transactions for specific shops (Hemkop and ICA chains) and all shops combined grouped and ordered by month)

-this category falls into COST OF LIVING; daily and house necessities, food&drink
-Hemkop and ICA represent two largest storechains in Sweden

CIGARETTES

(includes queries for total monthly transactions spent on buying cigarettes, again all costs and by specific Brand(Marlboro, Luckies - filtered by price) also grouped and ordered by month)

-this category falls into SMOKING HABITS

LUXURY

(last category sums up total transactions per month for non specified transactions like bars and restaurants)

Soon I had discovered that my longer queries were returning with invalid number errors. Upon closer inspection of the Data I have discovered couple of outliers. A few of the numbers over a thousand and including one thousand were displayed with a dot (.) behind a thousand number and a comma(,) for decimal spaces.

e.g. 1.000,00, to resolve the issue and remove the dot (.) from string we use following syntax

```
update croz
set amount = regexp_replace(amount, '\.', ''');
```

However I also found another solution.



SQL Queries

I had created a similar schema and changed transaction amount (Belopp) column from STRING to NUMBER.

```
CREATE TABLE croz_final(
    DATUM_PRIMJENE VARCHAR(45),
    VALUTA VARCHAR2(5),
    SREDNJI_ZA_DEVIZE NUMBER(10,5),
    TRANSAKTION VARCHAR2(45),
    BELOPP NUMBER(38,2)
);
```

Now I could insert all Data from original output excluding the "invalid numbers"

```
INSERT INTO croz
SELECT * FROM croz2
WHERE belopp NOT IN ('-1.582,83','1.000,00','2.000,00','3.000,00','4.000,00','5.000,00','-1.533,30','-1.407,35');
```

Result is a much cleaner database easier to work with without invalid numbers.

DATUM_PRIMJENE	VALUT	SREDNJI_ZA_DEVIZE	TRANSAKTION	BELOPP
2019-09-26	EUR	7,401888	Kortköp 190924 7 ELEVEN 4516110	-31,00
2019-09-26	EUR	7,401888	Kortköp 190924 BARAS BACKE	-29,00
2019-09-26	EUR	7,401888	Kortköp 190924 BARAS BACKE	-29,00
2019-09-26	EUR	7,401888	Kortköp 190924 BARAS BACKE	-29,00
2019-09-26	EUR	7,401888	Kortköp 190924 BARAS BACKE	-58,00
2019-09-26	EUR	7,401888	Kortköp 190924 HEMKÖP STOCKHOLM GUL	-140,59
2019-09-26	EUR	7,401888	Kortköp 190924 SHMI KIOSK	-59,00
2019-09-27	SEK	0,694617	Reservation Kortköp Baras Backe	-58,00
2019-09-27	SEK	0,694617	Reservation Kortköp HEMKÖP STOCKHOLM G	-67,79
2019-09-27	SEK	0,694617	Kortköp 190925 HEMKÖP STOCKHOLM GUL	-80,09
2019-09-27	SEK	0,694617	Kortköp 190925 HEMKÖP STOCKHOLM GUL	-114,09
2019-09-27	EUR	7,405313	Reservation Kortköp Baras Backe	-58,00
DATUM_PRIMJENE	VALUT	SREDNJI_ZA_DEVIZE	TRANSAKTION	BELOPP
2019-09-27	EUR	7,405313	Reservation Kortköp HEMKÖP STOCKHOLM G	-67,79
2019-09-27	EUR	7,405313	Kortköp 190925 HEMKÖP STOCKHOLM GUL	-80,09
2019-09-27	EUR	7,405313	Kortköp 190925 HEMKÖP STOCKHOLM GUL	-114,09
2019-09-28	SEK	0,691671	Reservation Kortköp HONEY HONEY	-54,00
2019-09-28	SEK	0,691671	Reservation Kortköp Direkten Tareqs To	-59,00
2019-09-28	SEK	0,691671	Reservation Kortköp HEMKÖP STOCKHOLM G	-158,12
2019-09-28	EUR	7,402264	Reservation Kortköp HONEY HONEY	-54,00
2019-09-28	EUR	7,402264	Reservation Kortköp Direkten Tareqs To	-59,00
2019-09-28	EUR	7,402264	Reservation Kortköp HEMKÖP STOCKHOLM G	-158,12

570 rows selected.

```
SUM(BELOPP)  
-----  
-58665,02
```

In practice we will use the first solution if we do not want to dismiss those transactions as irrelevant.

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

ANALYTICS

Here I will show some of the SQL queries I used to get the analytical data, export into Excel, provide you with links to uploaded reports and attach aditional report files. Also I will conduct a similar analytic in Jupyter Notebook Open source Tool and create extra Visualizations from Data insights.

Query used for first report (Bank statement for a single month) is simple. Also filtering by SEK is applied which gives us the option of converting Swedish Krona to HRK in Excel report

```
SELECT *  
FROM croz_final  
WHERE Datum_prmjene BETWEEN '2019-09-01' AND '2019-09-28'  
AND Valuta = 'SEK'  
;
```

Next query is used to report SUM of total transactions grouped and ordered by month filtered only by one currency so we dont end up with duplicate values.

```
SELECT ABS(SUM(Belopp)) AS Total,  
extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
FROM croz_final  
WHERE Valuta='SEK'  
GROUP BY extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
ORDER BY extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
;
```

Similar queries are written to retrieve all neccessary Analytical Data for the assignment.
e.g. total spent on shopping, luxuries, on cigarettes, and total spent in specific, shops, specific brand of cigarettes etc. Also what is important that I am using an ABSOLUTE function to convert all negative values to positive for easier visualization and converting column 'Datum_prmjene' from STRING to DATE.

```
SELECT ABS(SUM(Belopp)) AS Total_Shop,  
extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
FROM croz_final  
WHERE Valuta='SEK'  
AND Transaktion LIKE '%HEM%'  
OR Transaktion LIKE '%ICA%'  
GROUP BY extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
ORDER BY extract(month FROM TO_DATE(Datum_prmjene, 'YYYY-MM-DD'))  
;
```



PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

ANALYTICS

Subquery inside a query is needed to retrieve the SUM of monthly total spent on luxuries.

```
SELECT ABS(SUM(belopp)) FROM croz_final
WHERE transaktion NOT IN

(select transaktion from croz_final
where transaktion like '%ICA%'
or transaktion like '%HEM%'

AND belopp NOT IN('-57,00','-66,00')
AND valuta = 'SEK'
GROUP BY extract(month FROM TO_DATE(Datum_primjene, 'YYYY-MM-DD'))
ORDER BY extract(month FROM TO_DATE(Datum_primjene, 'YYYY-MM-DD'))
;
```

Retrieved results can now be easily exported from SQL Developer to Excel as separate tables for the purpose of creating a report. Before continuing to the next chapter which is not requested in the assignment itself (presents the alternative method done in Jupyter Notebook) please be advised to read the Report itself contained in the repository as a separate Excel file named Report.xls.

It contains two sheets;

- first one displaying the monthly transactions for September in different currencies
- second sheet displays pivot charts and actual conclusions drawn from discovering the clients habits insights from card payment transactions



PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

ZVONIMIR KLENOVIC

master engineer of graphic technology

820422-9150

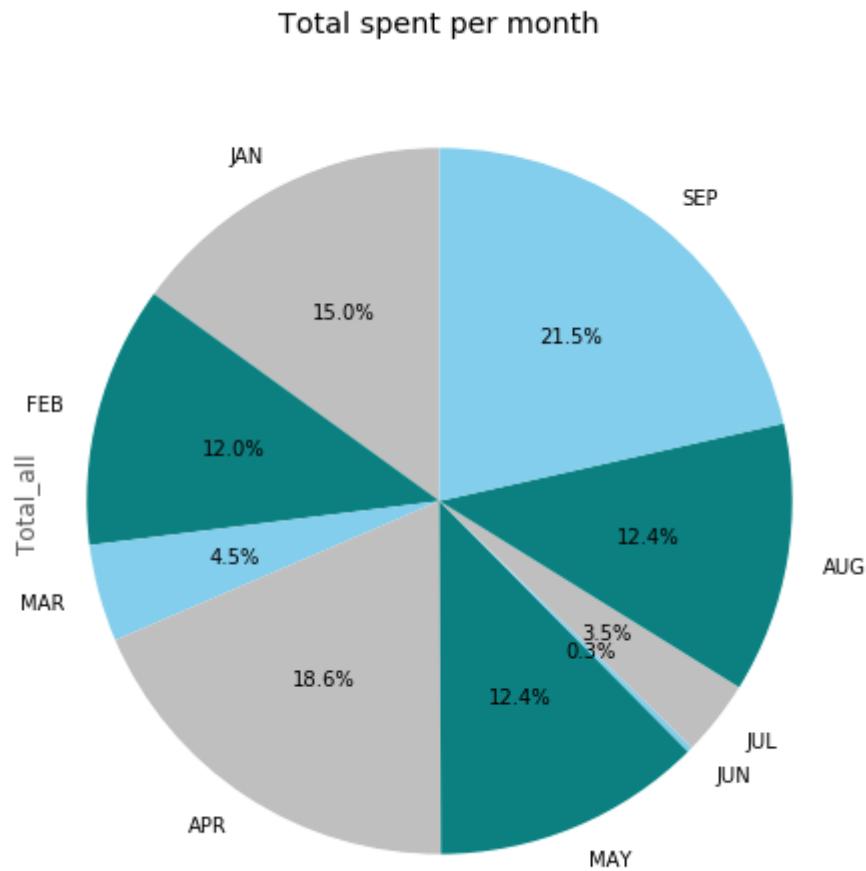
ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

JUPYTER NOTEBOOK

I have used Jupyter Notebook to show similar visualizations based on Analytical data and more. For this purpose I have created my own tables from exported excel data.

Please check out link to Jupyter Notebook for a complete overview of creating Dataframes from Excel Data and creating Visualizations.

https://nbviewer.jupyter.org/github/zvonimirklenovic/demo/blob/master/CROZ_Analitika.ipynb



PHONE

+385 914001982

email

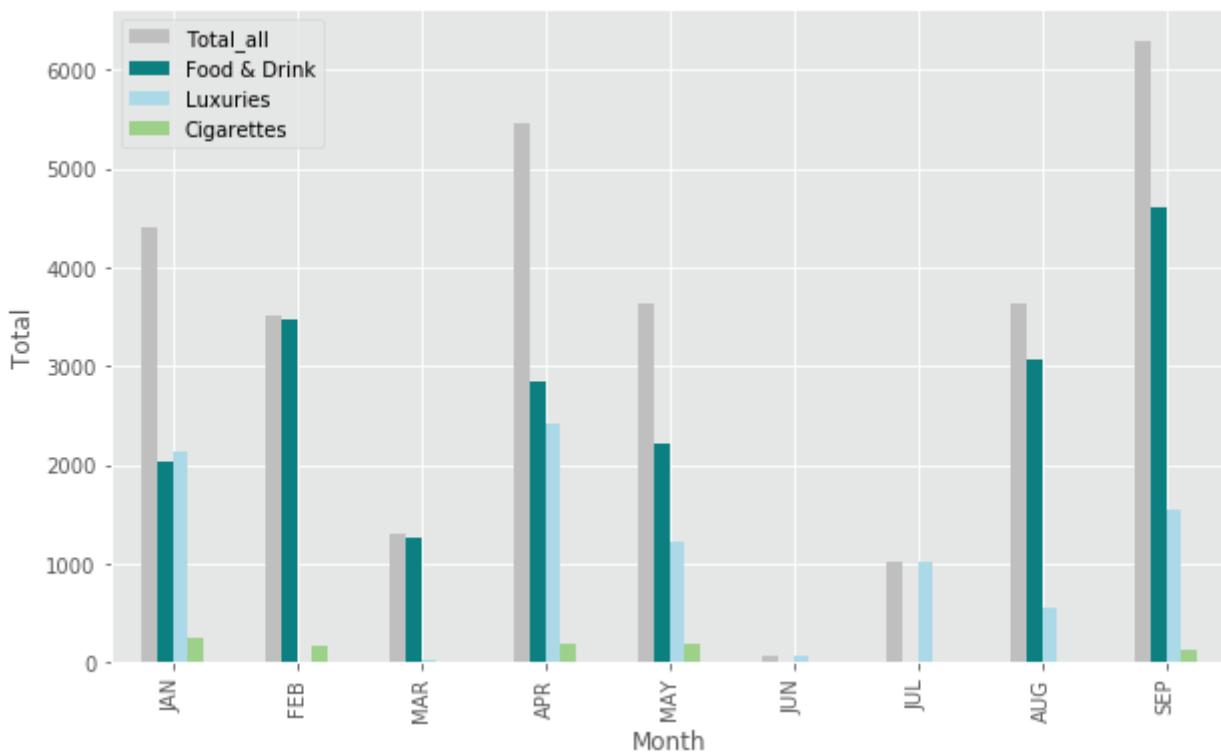
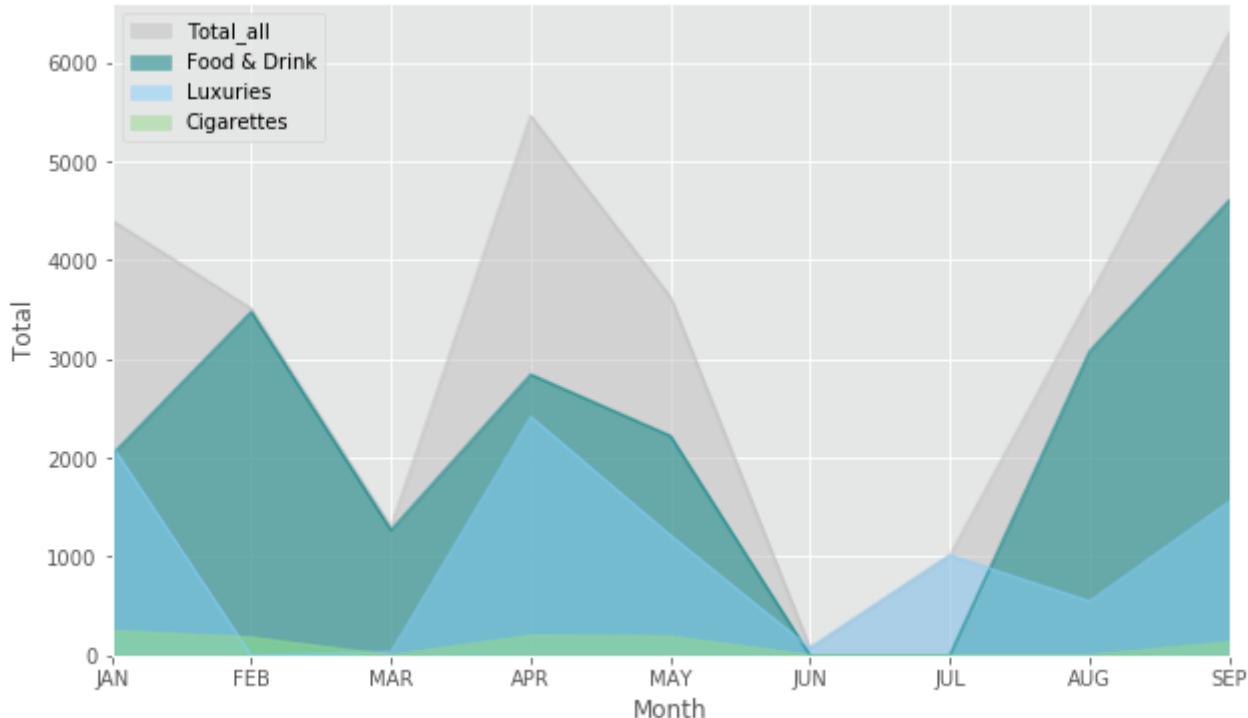
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**

master engineer of graphic technology

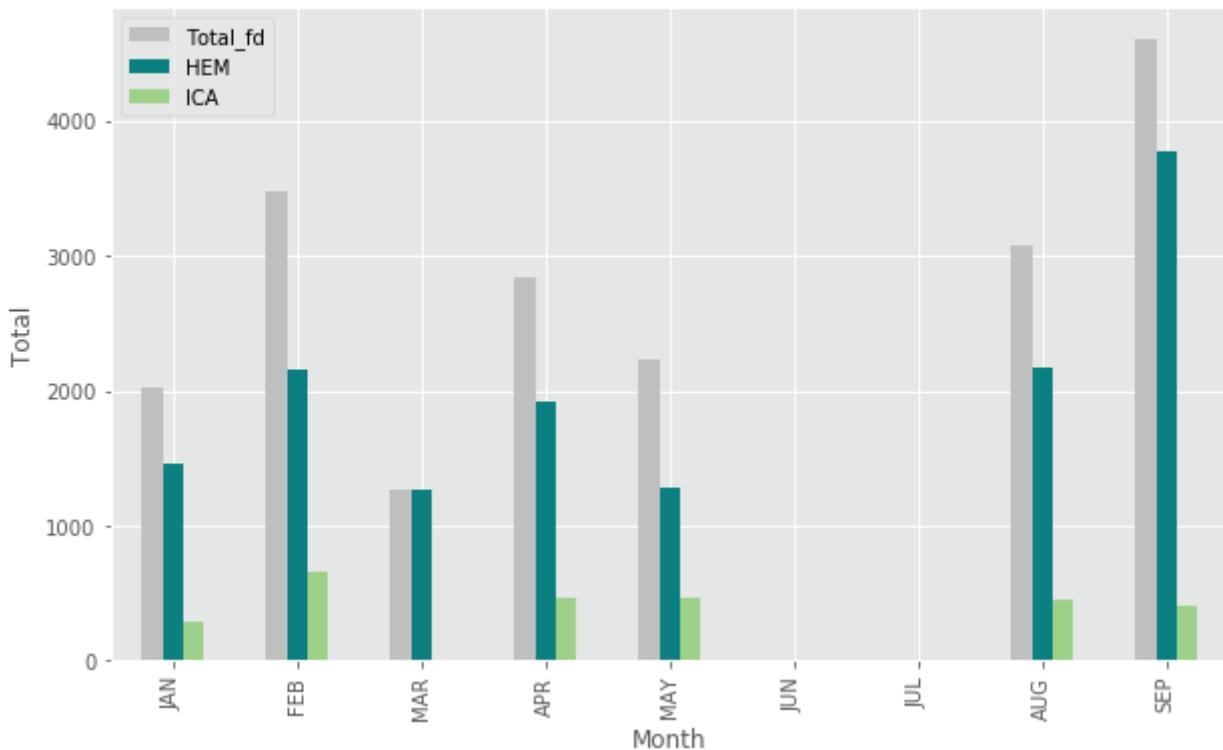
820422-9150**ADDRESS**Etsarvagen 10, Johanneshov 12143
Stockholm**JUPYTER NOTEBOOK**

Charts displaying Monthly spent per category. Some correlation is visible. It is described in report.



JUPYTER NOTEBOOK

In addition to the report if we look at the relationship between ICA and HEM we can notice that even though far less transactions are made in ICA there is a greater linear correlation and consistency in transactions. That might mean that ICA actually has a wider selection of products which can not be found in HEM but for that we would need to have insights into actual products linked to transactions.



Actual conclusion were drawn in Excel report. These Visualisations made in Jupyter Notebook are here to show an alternative approach and how the "Visual" part of visualizations actually looks great when matching the colours to the rest of presentation etc.

An original additional idea and how to use actual Analytical Data can be seen in the next section.

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

**ZVONIMIR
KLENOVIC**
master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

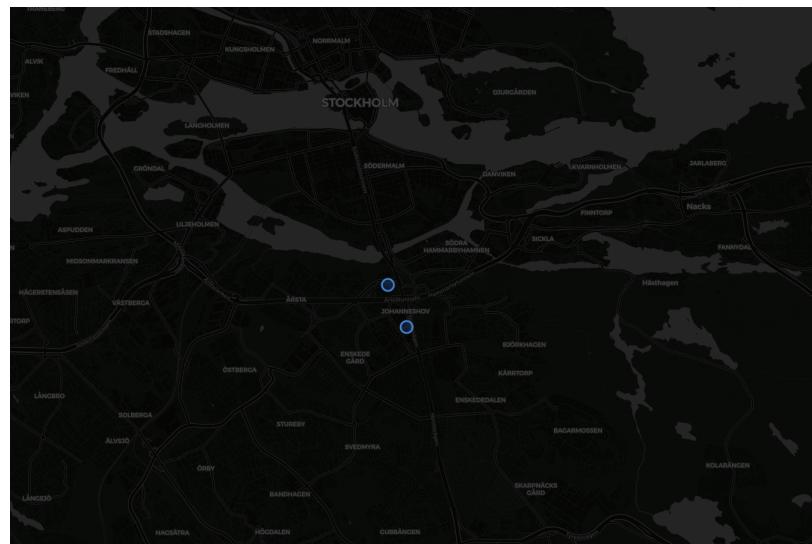
JUPYTER NOTEBOOK EXTRA

Addresses of two most frequent stores were taken and turned into geolocations using geopy library.

Hemkop Gullmarsplan
ICA Globen

(59.298476, 18.0790796)
(59.2923932, 18.084302)

Using the folium library I have superimposed Markers on an interactive Map on those locations. Based on the frequency of transactions in those stores we can conclude that our 'client' resides in that general area. This kind of information and visualisation can tell us more about the persons routine and whereabouts.



For full interactive map and code syntax please visit the notebook.

https://nbviewer.jupyter.org/github/zvonimirklenovic/demo/blob/master/CROZ_Extra.ipynb

PHONE
+385 914001982
email
zvonimir.klenovic@gmail.com

ZVONIMIR KLENOVIC

master engineer of graphic technology

820422-9150

ADDRESS
Etsarvagen 10, Johanneshov 12143
Stockholm

CONCLUSION

First of all I must say the assignment was not easy. I would like to thank you for the given opportunity and I would also like to thank myself for staying focused and dealing with issues in a calm and logical manner. Almost all of the technologies were completely new to me but I can honestly say that in this 2 weeks I spent working on the assignment I have learned more than in 6 months time of taking some online courses etc.

I got a chance to experience some of the daily routines and issues Data Engineers handle every day and based on that I have made certain conclusions;

NUMBER 01 - A lot of connectivity issues across platforms etc are mostly system related. In my opinion a lot of Data Engineers probably come from and have a strong System Engineer background.

NUMBER 02 - As far as actual ETL is concerned: "Schema is KING!", nothing beats a good schema it is the backbone of successful integration.

NUMBER 03 - You can never have enough knowledge of SQL . Queries get longer every day and so do subqueries :).

All I can say in my defence is that surely I have demonstrated I can indeed be a 'Fast learner' everyone is talking about. It felt great every time an issue got solved and that kept me going pretty much all the time. Data Engineering is really interesting and an important part of modern businesses. It is also in constant evolution as we are seeing newer technologies every day. One thing I have learned as an Engineer(of graph tech) is that an Engineers sole purpose is to improve. Well, I like to improve myself and help others do the same. This whole assignment has been very inspirational and I hope You have enjoyed reading this presentation as much as I did writing it. I am looking forward to hearing from You.

Sincerely,
Zvone