Cost of living Comparison
From Basel worker perspective
living in Basel, Freiburg or Colmar

Group 12:
Quynh Nhu Casati, Mario Corradini, Eitam Shafran

Data Collection, Integration and Pre-processing
Lecture Ramón Christen
Master in Applied Information and Data Science

Background

Freiburg

Colmar

Basel

Data

01
Colmar

Reppimo.com

36 apartments

	Surface [m2]	Rooms	Price [CHF/month]	Country	City
0	43.0	2.0	565	France	Colmar
1	84.0	4.0	997	France	Colmar
2	71.0	3.0	892	France	Colmar
3	60.0	3.0	593	France	Colmar
4	85.0	3.0	787	France	Colmar

02
Basel

homegate.ch

884 apartments

	Surface [m2]	Rooms	Price [CHF/month]	Country	City
0	99	3.5	3500	Switzerland	Basel
1	45	1	1450	Switzerland	Basel
2	48	2	1550	Switzerland	Basel
3	77	3	1830	Switzerland	Basel
4	54	1	1840	Switzerland	Basel

03
Freiburg

wg-gesucht.de

776 apartments

	Surface	Rooms	Price	Country	City
0	41	1.0	1074	Germany	Freiburg
1	18	1.0	824	Germany	Freiburg
2	46	1.0	1252	Germany	Freiburg
3	50	1.0	1252	Germany	Freiburg
4	26	1.0	918	Germany	Freiburg

Transform- explore and fix impurities of the data

Strings attached

43 m2
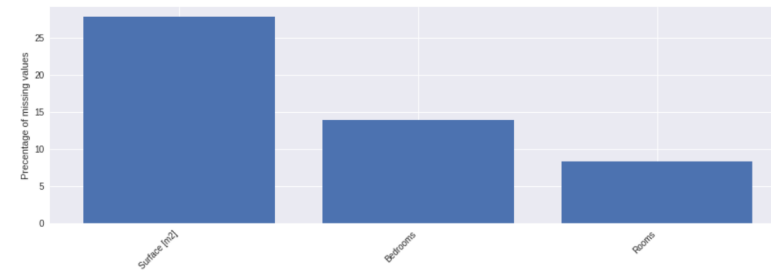
White spaces and unmatched currency

1 320 €/month

## Outliers
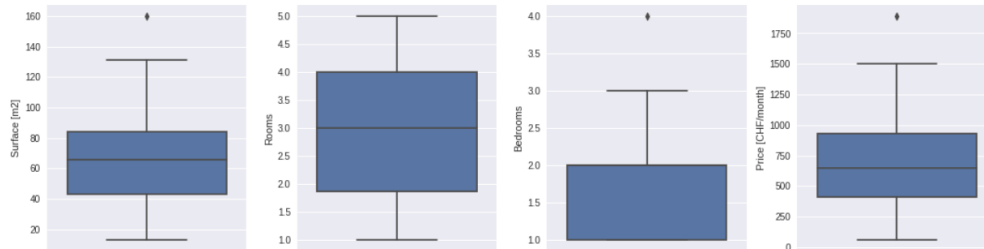
Missing values

Load

Creating new DataBase- Mysql connector

Upload data from python to data base- SQLAlchemy

Retrieving data from server in merged format

```
In [32]: ## Load tables from mariadb as 1 merged table
sql_statment = "SELECT * FROM \
    ( SELECT * FROM colmar_stage\
    UNION\
    SELECT * FROM freiburg_stage\
    UNION\
    SELECT * FROM basel_stage )a"

merged_tables = pd.read_sql(sql_statment, engine)
```
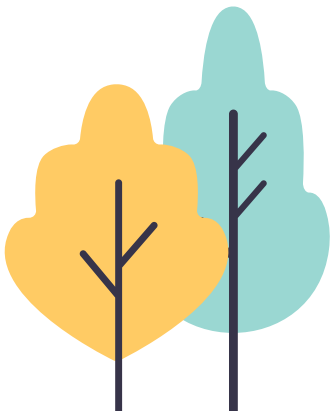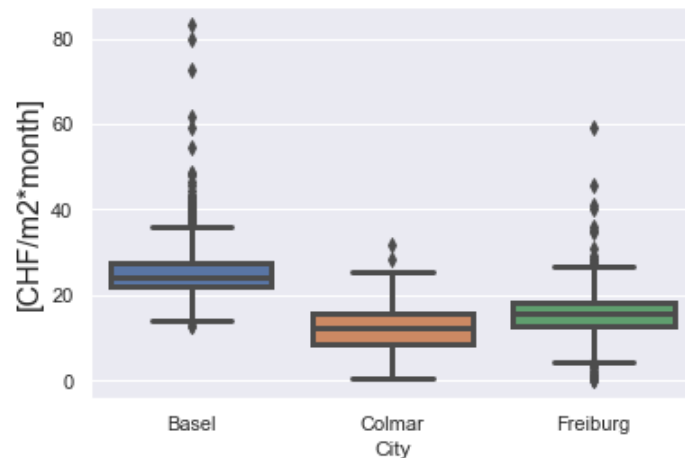
Answering questions – Question 1

What are the areas with the most expensive rents, considering different numbers of rooms?

	Surface	Rooms	Price	Country	City
0	41	1.0	1074	Germany	Freiburg
1	18	1.0	824	Germany	Freiburg
2	46	1.0	1252	Germany	Freiburg
3	50	1.0	1252	Germany	Freiburg
4	26	1.0	918	Germany	Freiburg

```python
# Rental per metter
df["rent_surface"] = df["Price"].div(df["Surface"])

# Different rental in general
sns.set(style="darkgrid")
sns.boxplot(x=df["City"], y=df["rent_surface"], linewidth=3)
plt.show()
```
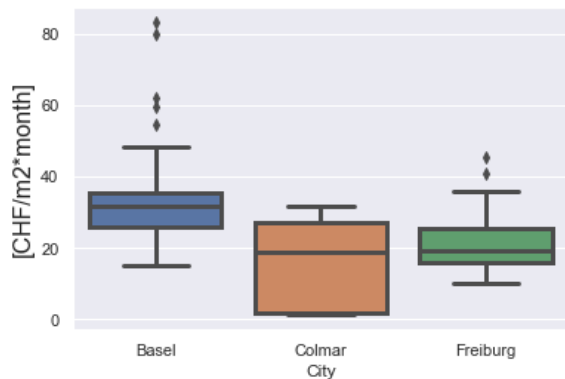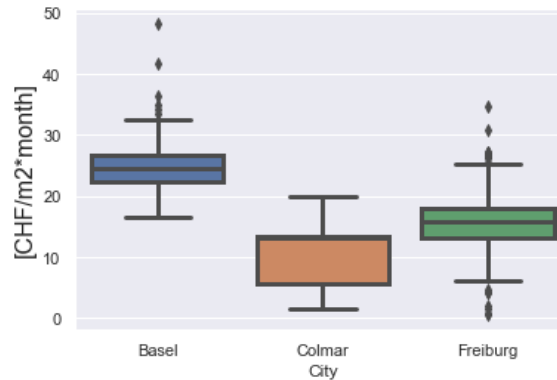
1. Rents in the three cities

Answering questions – Question 1

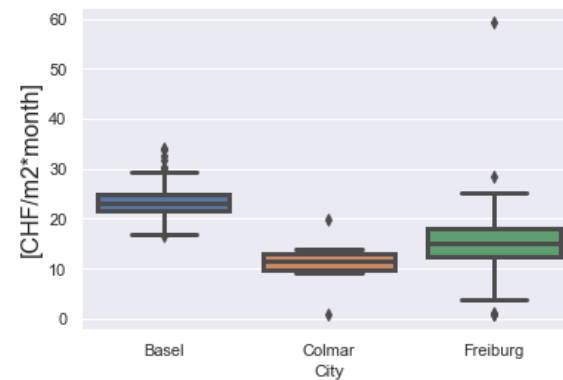**What are the areas with the most expensive rents, considering different numbers of rooms?**

2. Rents for 1-room-apartments	3. Rents for 2-room-apartments	4. Rents for 3-room-apartments


```
# one-way ANOVA test
_, pvalue = stats.f_oneway(df.loc[df["City"] == "Basel", "rent_surface"],
                           df.loc[df["City"] == "Freiburg", "rent_surface"],
                           df.loc[df["City"] == "Colmar", "rent_surface"])
print(pvalue)
```

4.456425029004657e-199

Answering questions – Question 2

How does the average price/m2 factor differ within each country, considering different apartment sizes?

Calculating the price per m2:

```
# First adding a column with the calculated price per m2, which I need to calculate for each entry
# Changing the column name because its easier to work with
df = df.rename(columns = {"Price [CHF/month]":"pricePerMonth"})
# Calculate the division of the mentioned column and add the result with the name pricePerm2 to the df
df["pricePerm2"] = df["pricePerMonth"] / df["Surface [m2]"]
```

Creating room size groups for France – the range is in between 9 – 250m2:

```
# We create five different categories from 9 to 250m2
# group_1 = [9-57m2], group_2 = [58-105m2], group_3 = [106-154m2], group_4 = [155-202m2], group_5 = [203-250m2]
# Grouping for France
group_1_FR = df[(df["Surface [m2]"] < 58) & (df["Country"] == "France")]
group_2_FR = df[(df["Surface [m2]"] > 57) & (df["Surface [m2]"] < 106) & (df["Country"] == "France")]
group_3_FR = df[(df["Surface [m2]"] > 105) & (df["Surface [m2]"] < 155) & (df["Country"] == "France")]
group_4_FR = df[(df["Surface [m2]"] > 154) & (df["Surface [m2]"] < 203) & (df["Country"] == "France")]
group_5_FR = df[(df["Surface [m2]"] > 202) & (df["Surface [m2]"] < 251) & (df["Country"] == "France")]
```

Calculating the mean for each group in France:

```
# Calculating the mean of each group from France
# There are no apartments for group 4 and 5 in Colmar
mean_1_FR = round(np.mean(group_1_FR["pricePerm2"]))
mean_2_FR = round(np.mean(group_2_FR["pricePerm2"]))
mean_3_FR = round(np.mean(group_3_FR["pricePerm2"]))
```
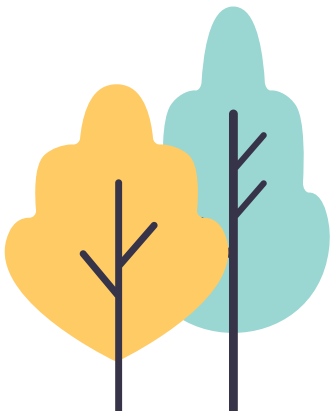
Answering questions – Question 2

How much does the average price/m2 factor differ within each country, considering different apartment sizes?

Table for France with all the mean values:

```
# Creating a table for France with all mean values
data_FR = [[mean_1_FR], [mean_2_FR], [mean_3_FR]]
df_mean_FR = pd.DataFrame(data_FR, columns=['France_mean'])
```
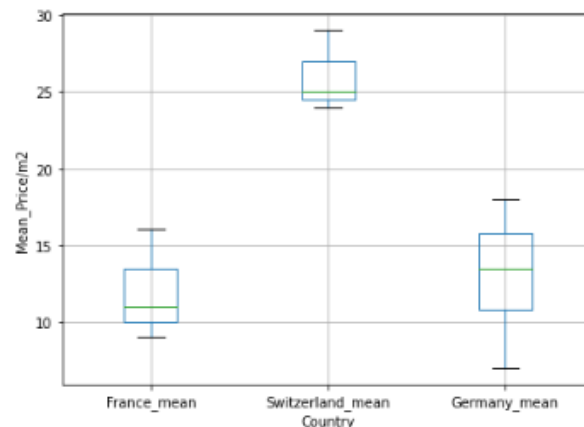
Concatenating all the means:

```
# Concatenating all three mean df's
frames = [df_mean_FR, df_mean_CH, df_mean_DE]
df_mean = pd.concat(frames)
```

Boxplot – comparison of the mean price/m2 factor :

```
# To make a fair comparison just the groups 1-3 can be compared.
df_mean.boxplot(figsize= (7,5))
plt.xlabel("Country")
plt.ylabel("Mean_Price/m2")
```
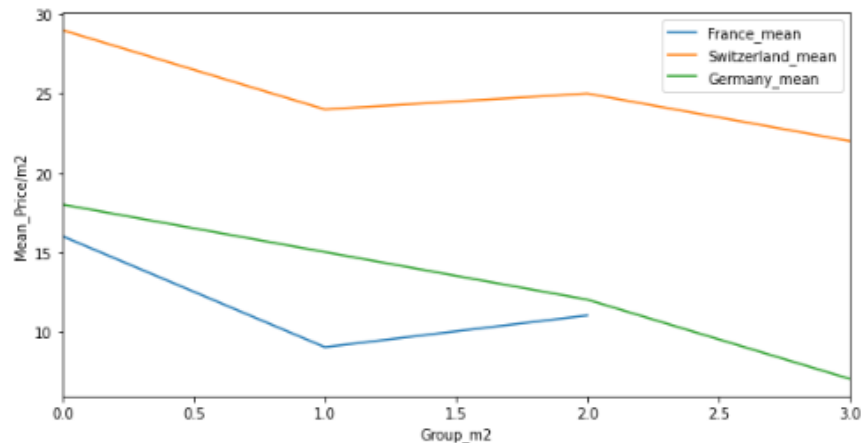
```
Text(0, 0.5, 'Mean_Price/m2')
```

Answering questions – Question 2

How much does the average price/m2 factor differ within each country, considering different apartment sizes?

Plot of the average Price/m2 for each country

Average Price/m2 factor tables

	price_m2_factor_CH_FR	Groups
0	1.81	9-57m2
1	2.67	58-105m2
2	2.27	106-154m2

	price_m2_factor_DE_FR	Groups
0	1.12	9-57m2
1	1.67	58-105m2
2	1.09	106-154m2

	price_m2_factor_CH_DE	Groups
0	1.61	9-57m2
1	1.60	58-105m2
2	2.08	106-154m2
3	3.14	155-202m2

Answering questions – Question 3

How would a person working in Basel rationally decide in which city he/she should rent an apartment of a desired size, also considering taxes and commuting costs?

Assumptions:
- A single person earning 100,000 CHF a year
- Standard vehicle: 9.4 liter/per100km
- Taxes: FR 0.41 CH 0.276 DE 0.416
- Fuel cost: 1.62 chf/liter

Answering questions – Question 3

```
## Calculating the net salary
net_salary["net_pay"] =  (100000/12)*(1-net_salary["taxes"])- net_salary["pricePerMonth"] -net_salary["Commuting"]*20
net_salary
```

	City	Surface_bins	pricePerMonth	taxes	Commuting	net_pay
0	Basel	(9.0, 57.0]	1172.635135	0.276	0.000000	4860.698198
1	Basel	(57.0, 105.0]	1782.077922	0.276	0.000000	4251.255411

Net salary after deduction by apartment size and city