Tentamen scriptprogramming GMI2BT

Regler och utförande

- Du har sex timmar på dig att utföra examinationen.
- Du får använda alla hjälpmedel **förutom** att kommunicera med andra personer.
- Identifiering:
  - Digital mötesplattform är Zoom.
  - Använd din DU-email och ditt fullständiga namn vid login på skolans digitala mötesplattformar: http://samtal.du.se.
  - Du ska via skolans digitala mötesplattformar ovan kunna visa giltig legitimation och ditt ansikte samt dela/streama din skärm till tentavakten
- Betyg och godkännande:
  - Det sätts inga poäng på varje enskild uppgift utan svårighetsnivån i samband med dina lösningar ger ett samlat totalbetyg.
  - För att få godkänt krävs att uppgiften är korrekt utförd där bedömningen görs utifrån lösningens funktionalitet, utförande och kvalité.
- Inlämning: Se till att din kod fungerar som den ska, packa sedan ner dina filer och skicka in i kursrummet där du hämtade tentamen. Innan du skickar in, testa att packa upp dina filer igen och kör programmet för att säkerställa att du skickar in rätt filer.

Lycka till!

Task 1: Access Netgear routers using their SOAP API

Background

The pyNetgear_Enhanced: https://pypi.org/project/pynetgear-enhanced/ library provides an easy to use Python API to control your Netgear router. It uses the SOAP-api https://en.wikipedia.org/wiki/SOAP on modern Netgear routers to communicate. It is built by reverse engineering the requests made by the NETGEAR Genie app. pyNetgear_Enhanced works with Python 3.

Task Information

In this task we are going to get mocked (fake/simulated) info from a modified pyNetgear_Enhanced library via the **stud_netgear_lib.py** file. You can inspect the faked module and the limited methods that it provides which return the same result on every call.

Use the **stud_netgear_start.py** file to begin creating your solution. There are several ways to solve the problem. One is using pandas with the `pd.DataFrame.from_dict()` and `dataframe.to_html()` methods which can be very useful.

a) Static web page

First task is to let Python write a static web page with router information, connected devices and network traffic stats like the attached **index.html** page file.

The Netgear routers connected devices should be able to sort by:

- IP-address,
- Name,

Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE

Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

- Type
- or not sorted at all when no parameter is given as argument

Traffic Volume in Megabytes | (Week or Month / Average) should be formatted according to the attached **index.html** page file. A correct HTML5 header and footer should be present so one can style the page if necessary.
If no argument were given the following output could be displayed:
Usage: netgear_stats.py <ip/name/type/none>

b) Dynamic web page
You should create a new python source file for this task.

Second task is similar to first the task and can share some code. In this case we generate the index.html page as a dynamic web page when a web client for example is browsing: http://localhost:5000/ip or every time a web client refreshes this URL.

In other words, the RESTful web service API should generate and show a web page with the routers current (real-time) status as the attached **index.html** page file.

The routers connected devices should be able to sort with the URL parameters <ip/name/type/none> according to:

- IP-address,
- Name
- Type
- or not sorted when no URL parameter is given

To begin testing the API one can use the VS Code extension mentioned in the example.http file.

Task 2: Real Estate Sold Price EDA and Visualization

Background
There are several datasets with sold housing/real estate (villa) data from Dalarna obtained from: https://www.booli.se/slutpriser/dalarnas+lan/322?sort=published. The datasets are attached to the exam in the booli folder with some PDF files containing examples dealing with real estate visualization.
The datasets are in JSON format and needs to be converted (and flattened) to CSV with the attached code (booli_flatten.py) so it can be used with pandas.

Task Information
Select one dataset. Work with it answering the questions using visualisation following the principles you learnt in the course.

Use at least the following columns from the dataset:
`'soldDate', 'soldPrice', 'constructionYear', 'latitude', 'longitude', 'municipalityName', 'livingArea', 'plotArea', 'rooms'`

Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

Answer the following questions:

- Group villas by rooms, constructionYear, municipalityName etc. and show info.
- How does m² livingArea, rooms, plotArea, location, etc. relate to the soldPrice?
- What is the number of sold objects for every price_m2 (price per m² = soldPrice/livingArea)?
- Which geographical area is the most expensive? Which geographical area is the cheapest?
- What correlation do we have between the variables?

The following example is done with sold villas in whole Dalarna county. **But you can select another dataset, e.g. the county you live in.**

Example EDA

Please note that the example is flattened with cherrypicker (booli_flatten.py).

1. Show basic info about the dataset. Drop NaN (null/Na values) and add a price_m2 (soldPrice per m² = soldPrice/livingArea) column.

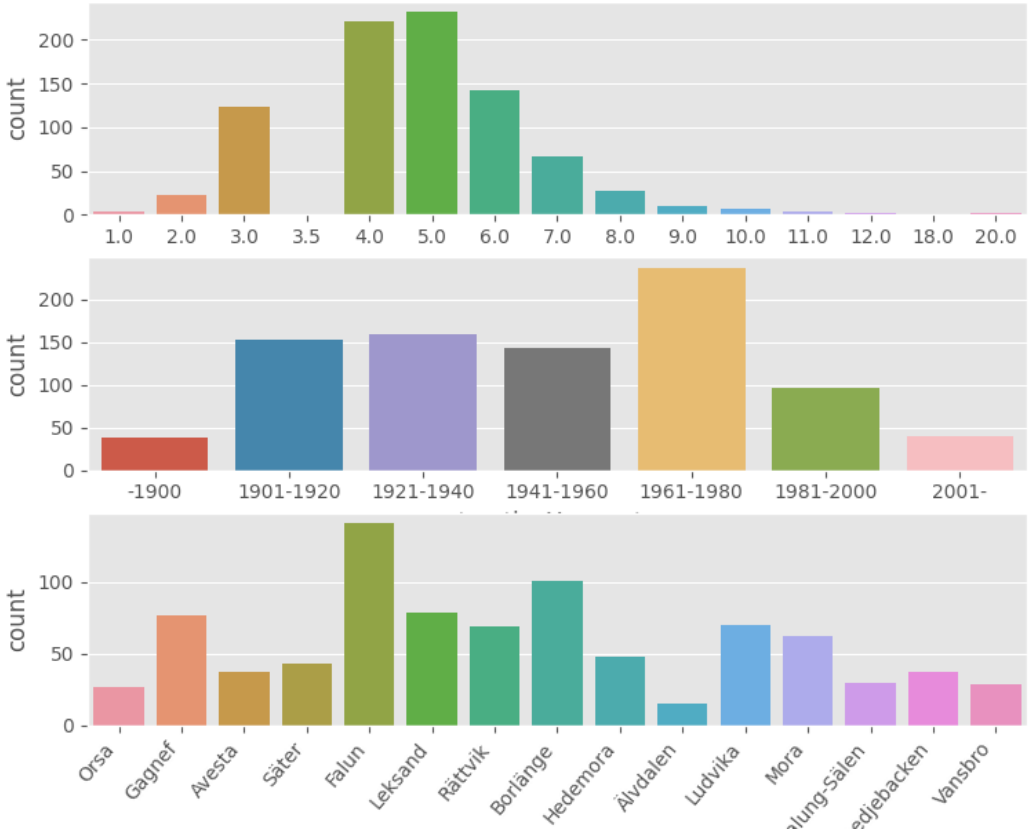
2. Group by municipalityName and visualize mean values for Dalarna state.

```
df_municipality_mean:
                soldPrice  constructionYear  latitude  longitude  livingArea  plotArea  rooms  price_m2
municipalityName
Avesta           1.97e+06           1946.16     60.17      16.27      131.46  10479.70   4.81  14320.22
Borlänge         2.69e+06           1958.44     60.48      15.42      128.08   1681.33   4.70  21290.01
Falun            3.27e+06           1949.89     60.64      15.68      140.91   2196.30   5.35  24091.43
Gagnef           1.83e+06           1945.83     60.54      15.06      132.36   1973.26   4.58  15197.31
Hedemora         1.51e+06           1949.38     60.32      15.99      131.92   1891.60   5.04  12011.63
Leksand          2.46e+06           1941.92     60.73      15.02      134.12   2691.10   4.96  19356.25
Ludvika          1.63e+06           1950.96     60.16      15.08      133.51   1942.46   5.20  12716.48
Malung-Sälen     1.96e+06           1960.03     60.94      13.42      140.03   3518.40   5.17  14903.67
Mora             2.00e+06           1953.94     61.00      14.52      123.22   2315.15   4.82  16518.66
Orsa             1.64e+06           1952.26     61.13      14.66      154.63  14572.78   5.15  14872.08
Rättvik          1.94e+06           1942.22     60.97      15.13      132.17   2750.88   4.81  15902.16
Smedjebacken     1.64e+06           1952.49     60.12      15.40      122.57   2588.89   4.95  14537.59
Säter            1.92e+06           1936.95     60.38      15.71      125.19   1875.58   4.86  15742.86
Vansbro          6.76e+05           1943.34     60.51      14.29      120.12   2439.03   4.41   6326.37
Älvdalen         1.10e+06           1942.47     61.58      13.33      121.73   3216.13   5.13   9526.74
```

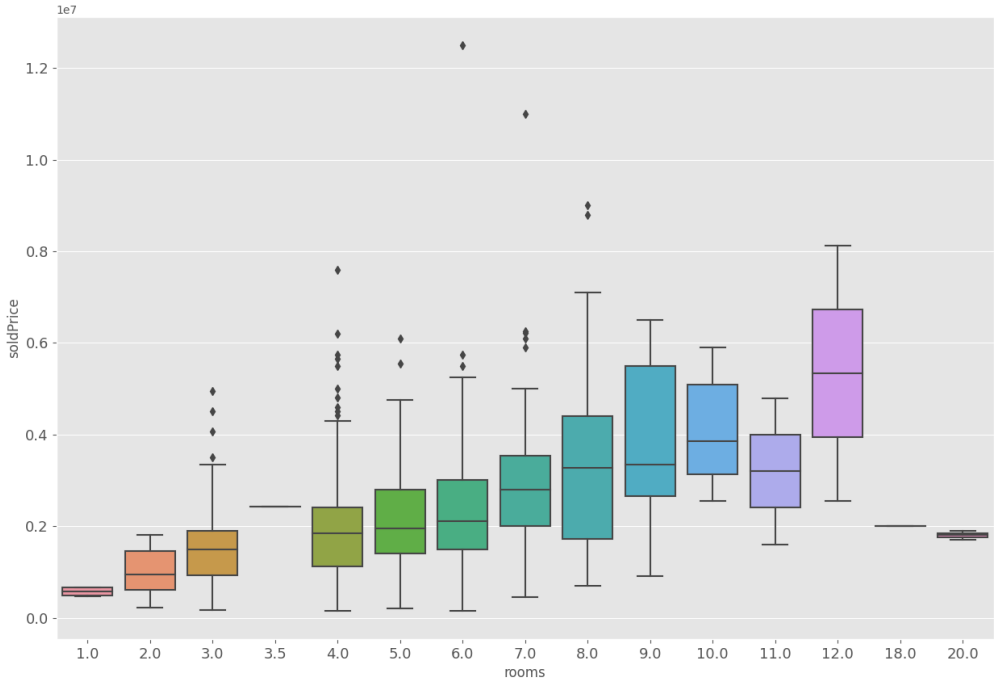
and/or group by: a) rooms, b) constructionYear or c) municipalityName. Visualize the DF using bar plots for every outcome as shown below.

Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE

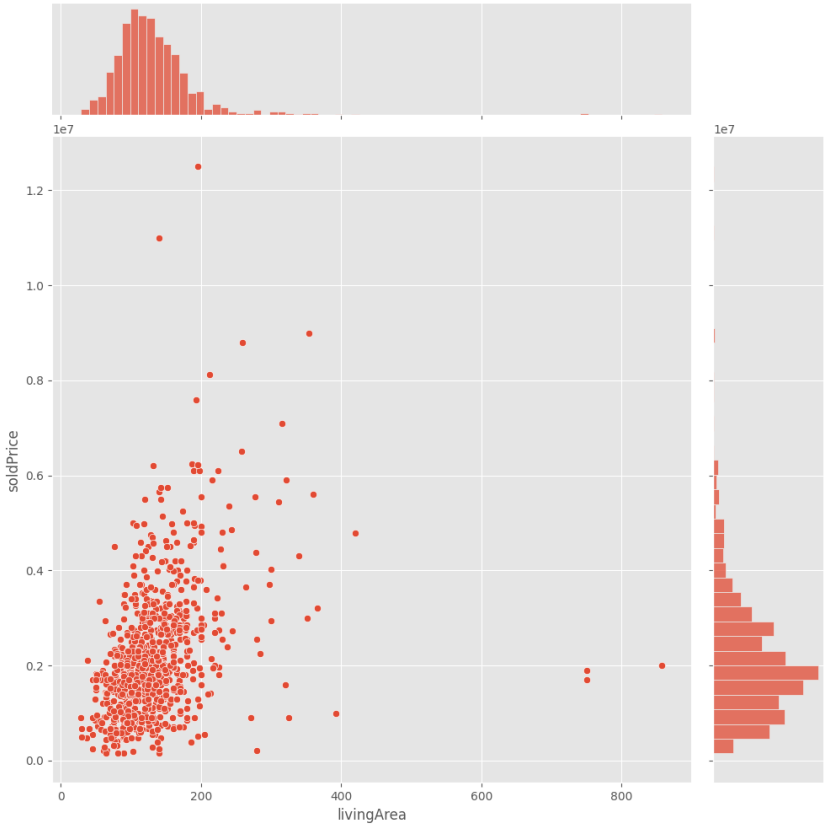
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/



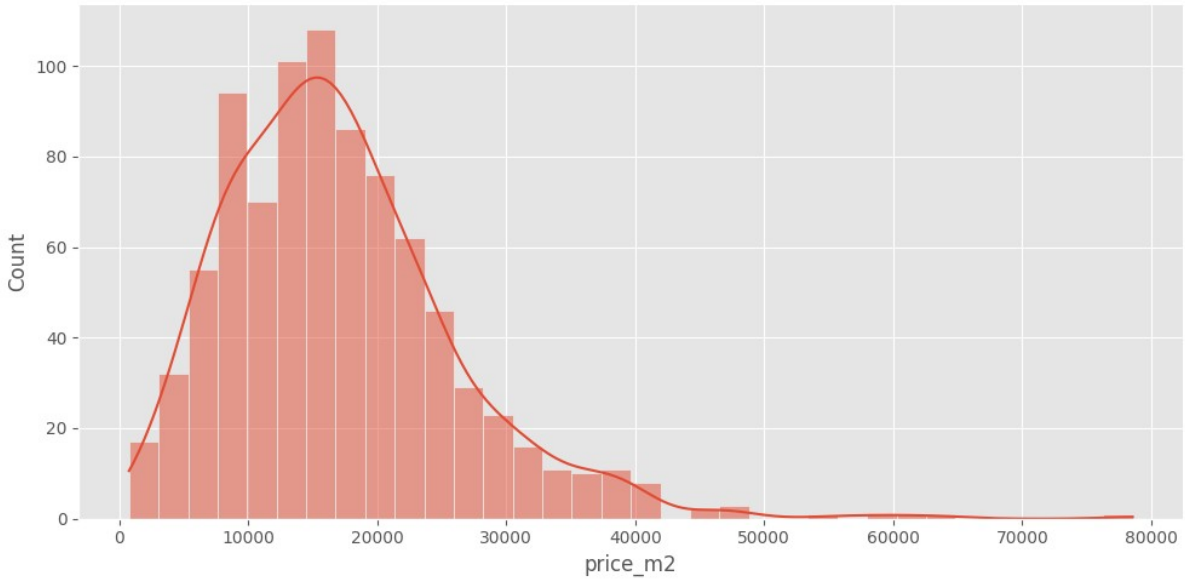
3. Boxplot the number of rooms vs. the soldPrice. Remove outliers if needed. Do the same for plotArea and LivingArea. An example is shown in the figure below.



4. How does livingArea m² relate to soldPrice? Outliers was not removed here. Reflect about your thoughts if they should be removed. Why and how?



5. Plot the
number of sold objects per price_m2 using bar plots and a connecting line at top of every bar. You can adjust the bar width as well. Note: price_m2 (soldPrice per m²) = soldPrice/livingArea.

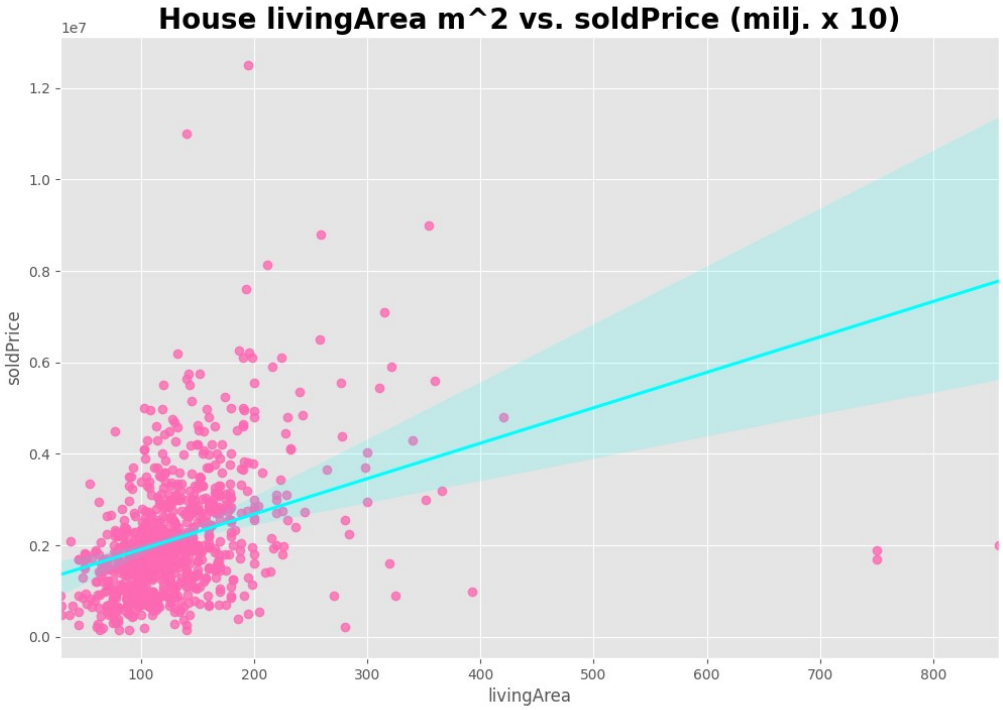


Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

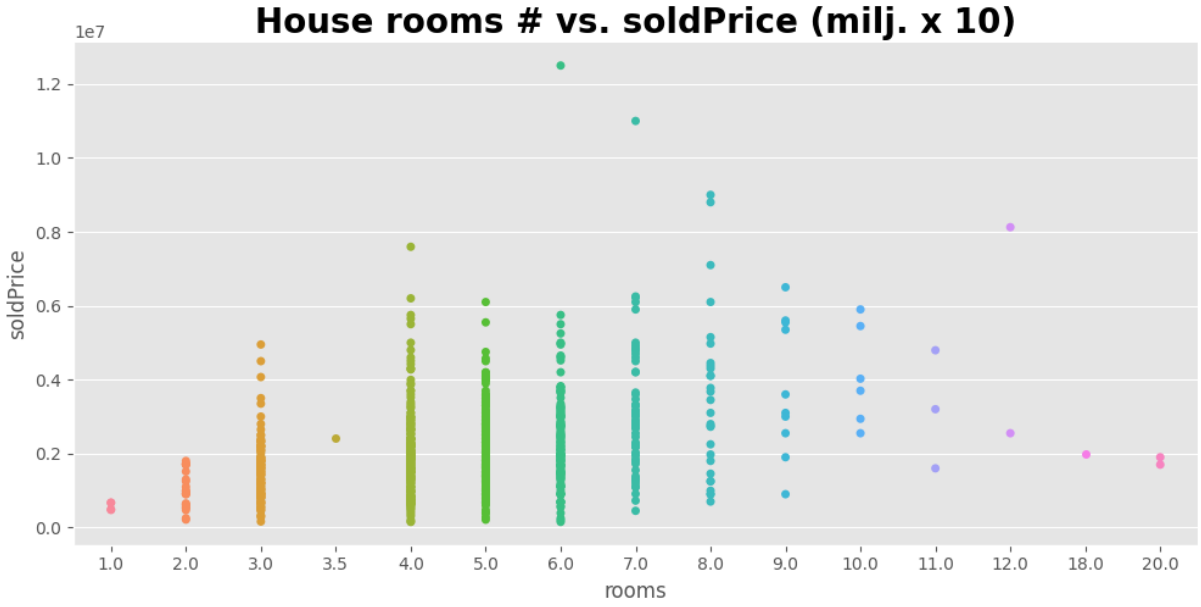
6. Visualize the correlation between three variables soldPrice, livingArea, and price_m2. Note the correlation between the variable to itself will be plotted as histograms and to other variables should be in form of scatter plot.

7. Plot the livingArea in m² vs soldPrice? Add a regression line to the plot. How does the two variables relate to each other? Will the results different after removing the outliers?

Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

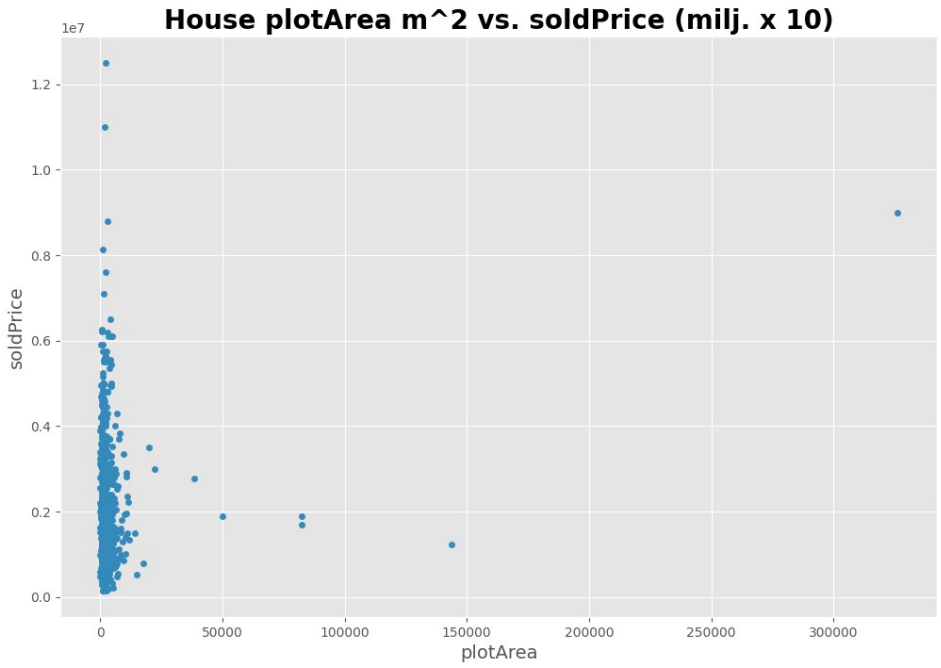


8. Plot the number of rooms (x) vs soldPrice (y). Is there any relation between number of rooms and soldPrice? Elaborate your answers.

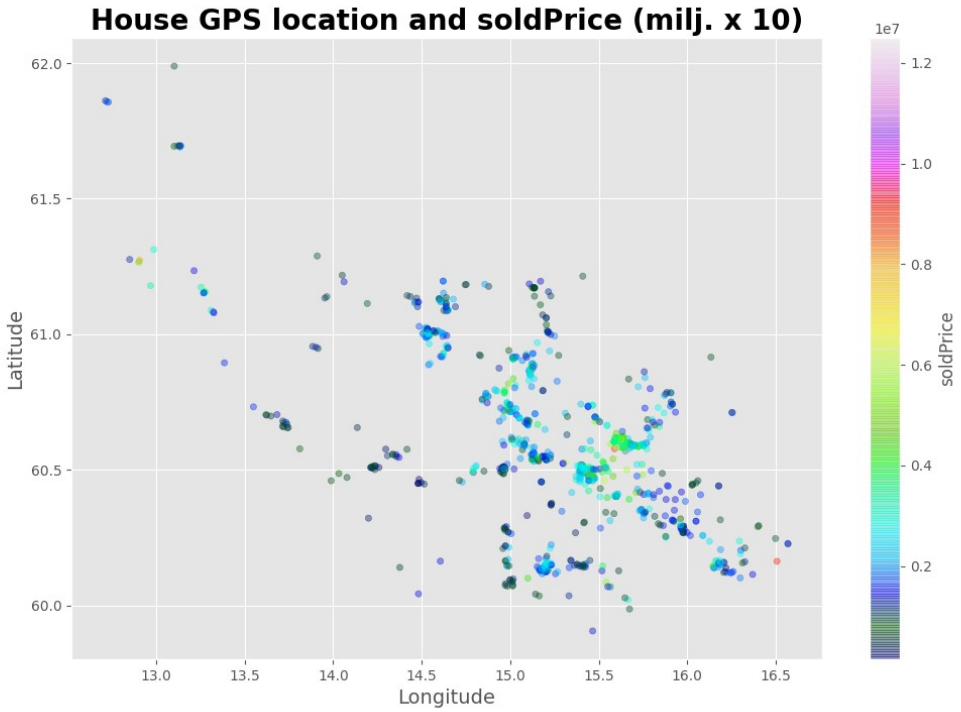


Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

9. Visualise the plotArea m² vs soldPrice using scatter plot. Describe your understanding from this figure with and without removing the outliers.
An example is shown in the below figure (Outliers was not removed).



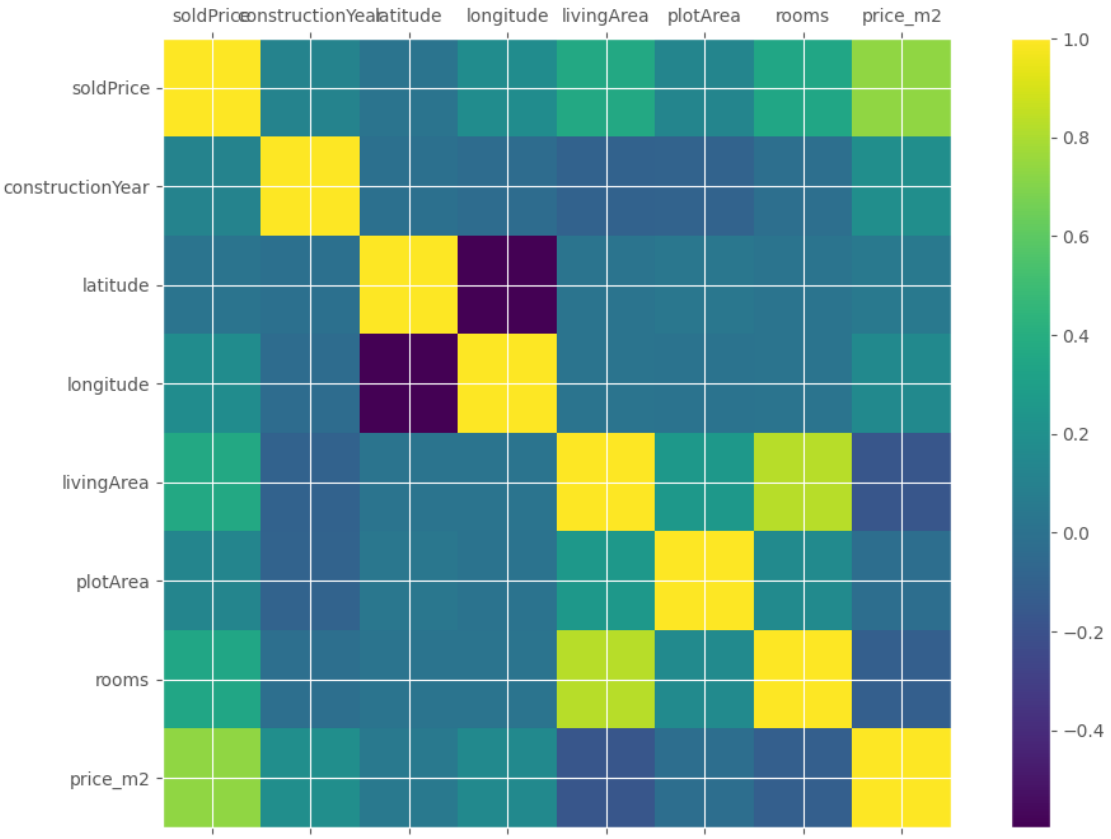
10. Assuming Longitude = x and Latitude = y, after plotting the x and y (Dalarna county) and colouring the points based on sold Price, describe your analysis for how does the longitude and latitude impact soldPrice? Can you identify an area for expensive and cheap prices?



Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE
Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/

You can use: https://www.hitta.se/kartan!~60.45520,15.61975,8.528029689019846z/tr!i=UQ5oSxpZ/GPS!l=60.48045:15.43156?usergeo=1 and set Koordinater to WGS84 DD (lat, long) to locate the villas.

11. Look into the correlation between only the numerical variables/columns (similar to below plot). Show the correlations. Note that price_m2 should not be included since it is derived from soldPrice and livingArea.



Högskolan Dalarna
Röda vägen 3
781 88 BORLÄNGE

Telefon: 023-77 80 00
Telefax: 023-77 80 50
URL: http://www.du.se/