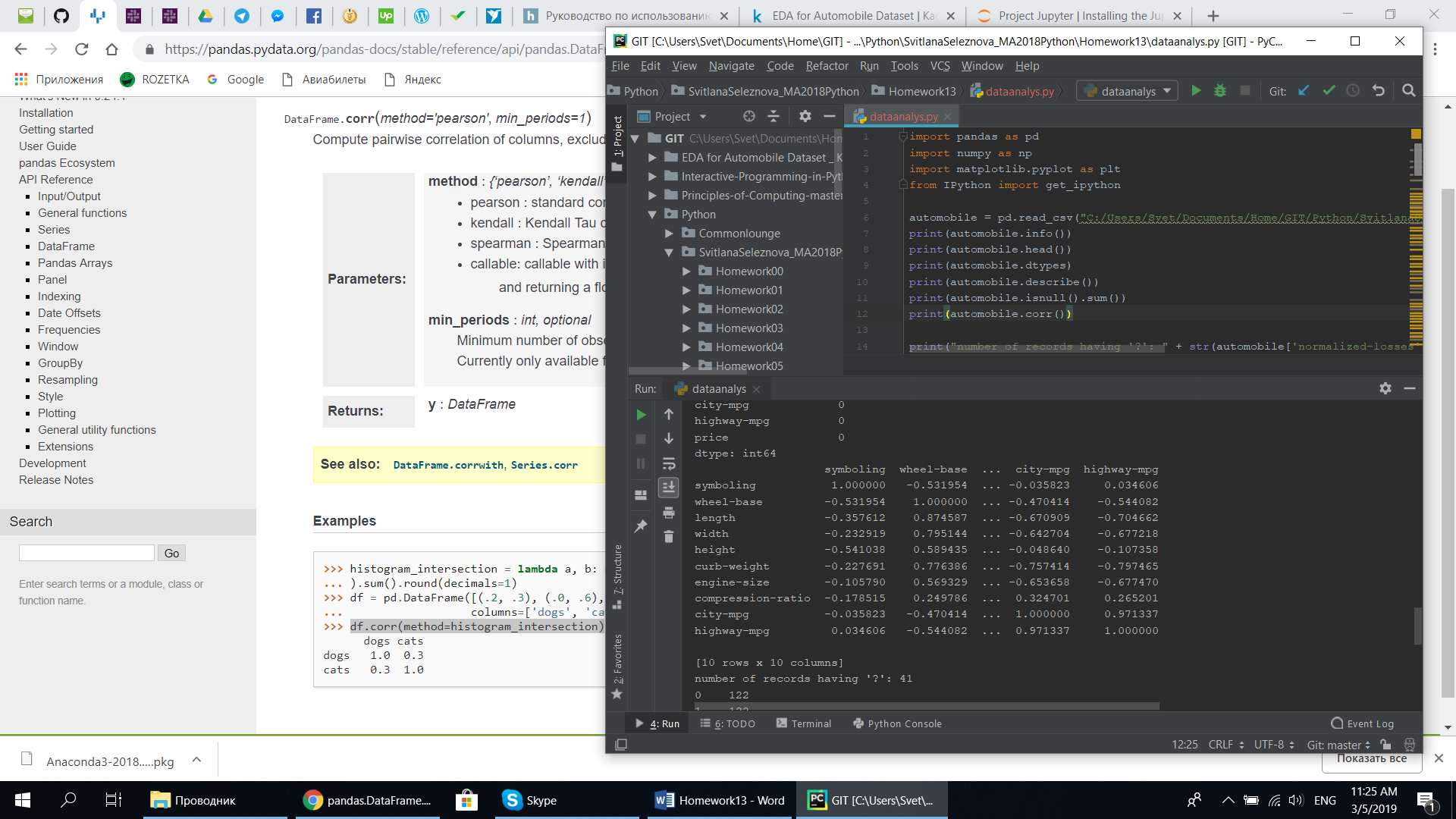
|  |  |
| --- | --- |
| **Numerical attributes** | **Categorical attributes** |
| Symbolling  normalized-losses  num-of-doors  wheel-base  length  width  height  curb-weight  num-of-cylinders  engine-size  bore  stroke  compression-ratio  horsepower  peak-rpm  city-mpg  highway-mpg  price | make  fuel-type  aspiration  body-style  drive-wheels  engine-location  engine-type  fuel-system |

2. А) Перевіряємо чи є нульові значення. В даному випадку нульових значень немає.

Б) Необхідно нечислові атрибути, які являються числовими, конвертувати в числові (normalized-losses, price, horsepower, bore, peak-rpm, num-of-doors).

В) Перевіряємо кожне поле на відсутність даних - ?.

* в колонці «normalized-losses» є 41 значення «?» - ці значення необхідно конвертувати в числове значення як середнє значення
* в колонці «price» необхідно знайти та конвертувати також нечислові 4 значення «?» в числові
* в колонці «horsepower» необхідно знайти та конвертувати також нечислові значення «?» в числові
* в колонці «bore» необхідно знайти та конвертувати також нечислові значення «?» в числові або прирівняти до 0
* в колонці «stroke» необхідно знайти та конвертувати також нечислові значення «?» в числові або прирівняти до 0
* в колонці «peak-rpm» необхідно знайти та конвертувати також нечислові значення «?» в числові або прирівняти до 0
* в колонці «num-of-doors» є значення «?» - цими значеннями можна знехтувати і очистити їх.



1. - Price більше співвідноситься з engine size і curb weight

* Curb weight в основному співвідноситься з engine size, length, width та wheel based, оскільки вони збільшують вагу автомобіля
* Wheel base сильно корелюється з length і width автомобіля
* Symboling і normalized-losses більш співвідносяться, ніж інші поля
* Чим більший розмір двигуна, тим дорожче ціна (колінеарні атрибути - engine-size, price)