**УДК 004.94, 519.2**

**Опыт эконометрического моделирования стоимости квартир на рынке недвижимости с использованием пакета статистического анализа Python**

**Аннотация.** Сегодня, масштабному применению эконометрических методов может мешать высокая стоимость коммерческих пакетов статистического анализа. Выходом из сложившейся ситуации может быть использование открытого программного обеспечения, удачным примером которого является пакет статистического анализа языка Python. Python - это современный язык программирования, востребованный и популярный в мировой научной среде. В статье демонстрируются новые возможности ценообразования на рынке недвижимости.

**Ключевые слова:** эконометрика, модель, Python, статистика, уравнение регрессии.

Специфической особенностью деятельности экономиста является работа в условиях недостатка информации и неполноты исходных данных для моделирования. Анализ такой информации требует использования специальных методов, которые составляют один из аспектов эконометрики. Одной из центральных проблем эконометрики является построение эконометрической модели и определение возможности ее использования для описания, анализа, прогнозирования реальных экономических процессов.  Сегодня, применению эконометрических методов может мешать высокая стоимость коммерческих пакетов статистического анализа. Выходом из сложившейся ситуации может быть использование открытого программного обеспечения, удачным примером которого является пакет статистического анализа языка Python. Python - это современный язык программирования, востребованный и популярный в мировой научной среде. В настоящей работе демонстрируются новые возможности ценообразования на рынке недвижимости.

Язык программирования Python - это мощный высокоуровневый кроссплатформенный язык. Он поддерживает объектно-ориентированное программирование, и в последнее время стал серьезной альтернативой таким языкам программирования как C++. В отличие от MATLAB, язык Python изначально не заточен под научные вычисления.

Библиотека Pandas языка Python позволяет с легкостью манипулировать исходными данными и анализировать их.  В принципе, библиотека Pandas построена на еще одной замечательной библиотеки в python 3 – Numpy. Использование пакета Pandas дает широкие возможности при работе с электронными таблицами. Данный пакет делает Python мощным инструментом для анализа данных. Пакет имеет возможность строить сводные таблицы, выполнять группировки, предоставляет удобный доступ к табличным данным, а при наличии пакета [matplotlib](http://matplotlib.org/) дает возможность визуализировать графики на полученных данных. Ниже будут показаны основы работы с пакетом, такие как загрузка данных, обращение к полям, фильтрация и построение сводных.

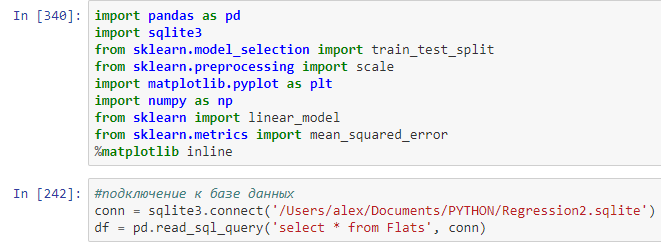
В статье предлагается использовать широкие возможности языка Python, в качестве инструмента обеспечивающего высокую производительность и точность при использовании эконометрических и статистических методов анализа данных.

В качестве примера использованы рассмотрены цены на квартиры в Москве и Московской области.

В результате данного анализа были выделены некоторые факторы, которые наиболее ощутимо влияют на стоимость квартиры, разработана многофакторная математическая модель описывающая процесс ценообразования на рынке недвижимости, получены уравнение регрессии и матрицы корреляции переменных, построены графики влияния исследуемых факторов на стоимости квартир.

В работе рассматриваются широкие возможности открытого и свободного программного обеспечения – FLOSS (Free\Libre and Open Sourse Software). Исследуются особенности ценообразования.

Объектом исследования является рынок недвижимости, целью – выявление критериев оценки и уровня их влияния на стоимость квартиры. Данные для проведения исследования были получены с web-сайта Cian.ru– крупнейший интернет-сервис для покупателей и арендаторов жилья, самая большая и актуальная база объявлений, регулярно обновляемая за счет запуска новых сервисов и расширения географии представленных объектов.  
Выбор сайта объясняется, во-первых, тем, что сайт имеет достаточно большую базу предложений, во-вторых, по каждому продаваемому объекту недвижимости в базе имеется подробная информация о его характеристиках.  
Отметим, что более половины жителей двух столиц подбирают жилье на Cian.ru. За последние два года эта часть аудитории выросла в два раза и сейчас достигает 3 млн покупателей в месяц. По данным исследований в Московском регионе, **60% всех звонков по продажам квартир в Москве и Московской области раздается благодаря объявлениям на Cian.ru.**



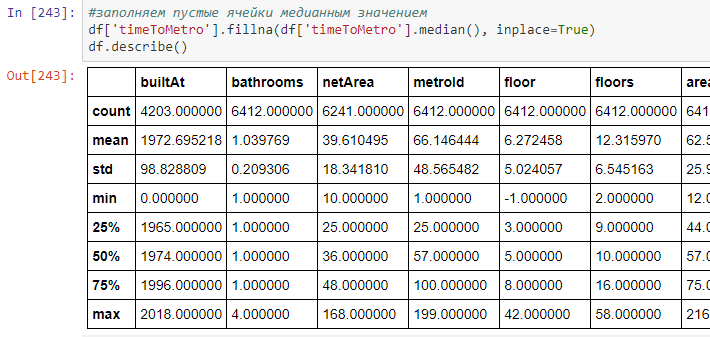
О каждой квартире в извлеченной выборке имеется следующая информация: стоимость (**price),** год постройки **(builtAt)**, количество комнат (**rooms**), площадь **(area)**, жилая площадь (**netArea),** этаж (**floor)**, этажей в доме (**floors)**, количество лифтов (**elevators**), балкон **balcony**, лоджия (**loggia),** санузел (**bathrooms)**, метро **(metroId)**, расстояние до метро (**timeToMetro)**, статус помещения, окна, и т.д.

Для построения модели использованы такие модули как: pandas, ведущая библиотека для визуализации данных в statsmodels.formula.api, бибилиотека двумерной графики matplotlib.pyplot.

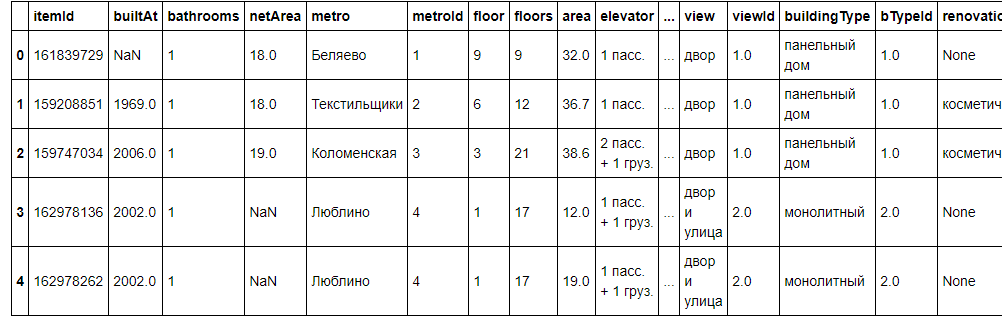
По теме исследования имеются работы, в которых не представлены современные эконометрические модели, которые могут включать не только числовые параметры.

В полученной таблице данных имеются пустые ячейки, для получения адекватного инструмента моделирования и прогнозирования заполним их медианными значениями соответствующих характеристик.

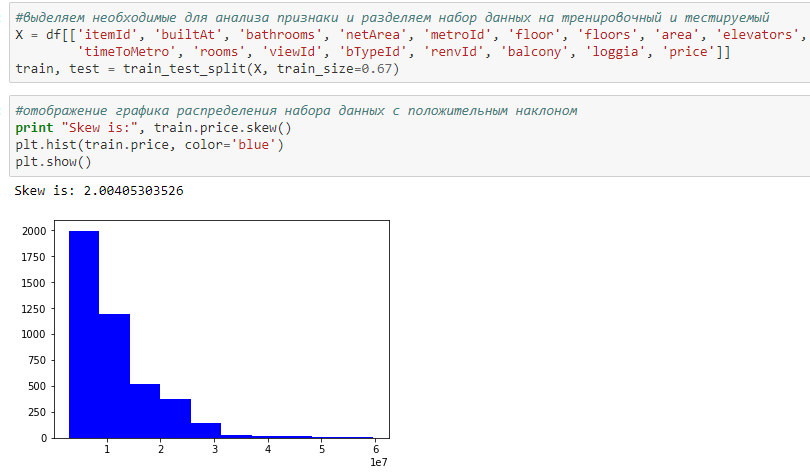
Получены следующая статистическая информация о выборке



Ниже приведены первые несколько строк таблицы данных, по которым хорошо видно, что пустые ячейки отображаются как неопределенные данные типа NaN, таких ячеек довольно много, поэтому пренебрегая ими можно получить грубую и неадекватную модель.

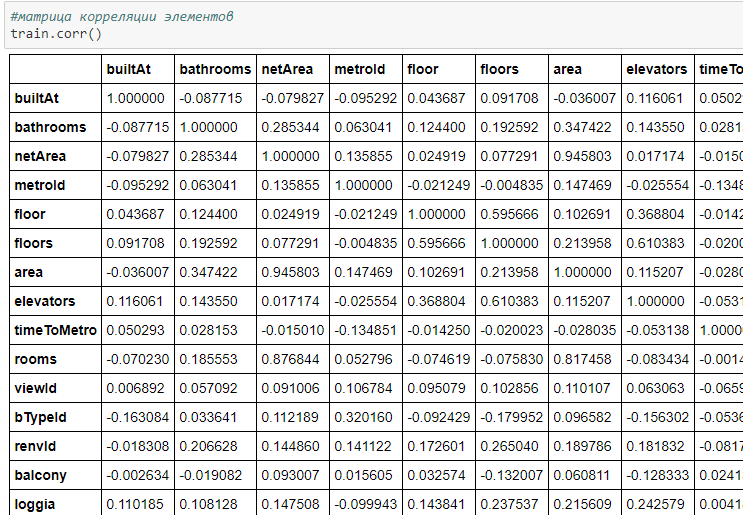


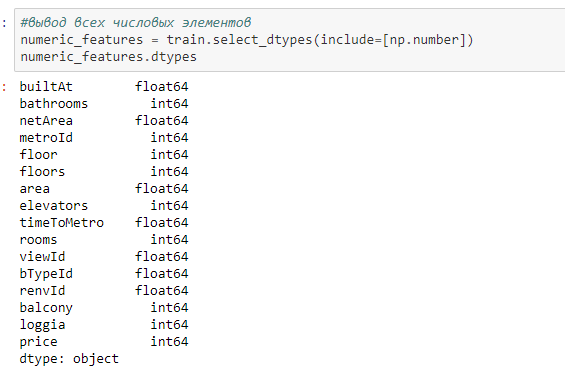
Для того, чтобы оценить результаты полученной модели, необходимо разделить имеющуюся базу данных объектов недвижимости разделить на тренировочные (данные по которым модель обучается) и тестируемые (для оценки точности прогноза, полученного в результате применения модели)

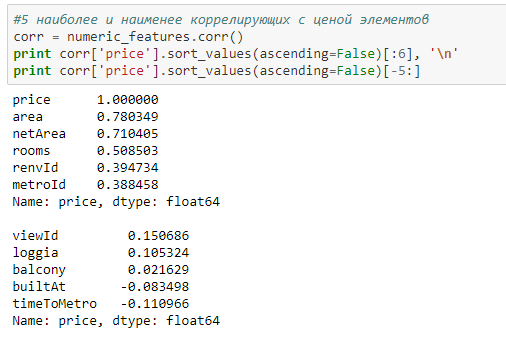


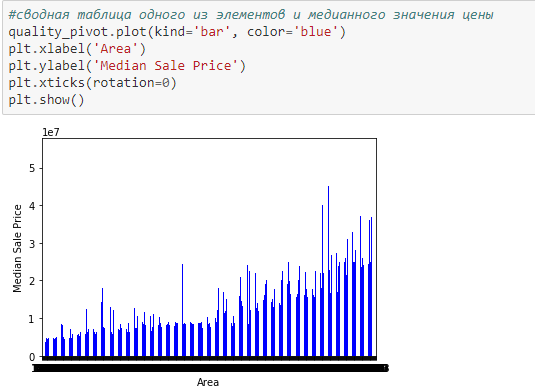
Проведен корреляционный анализ извлеченной из базы выборки, результаты приведены в таблице.

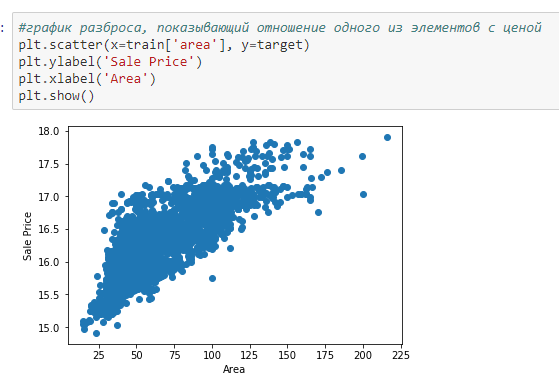
Рисунок 3. Матрица корреляции переменных для квартир.

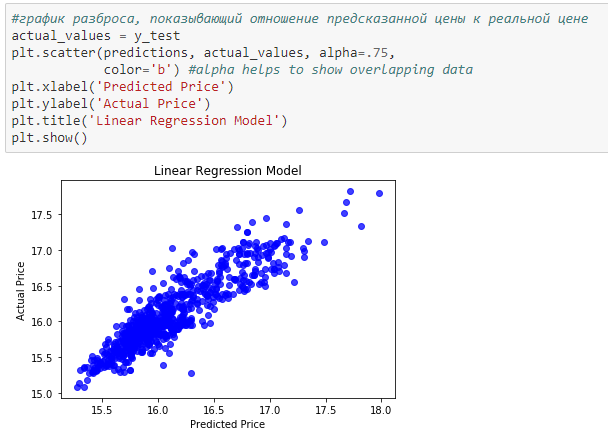












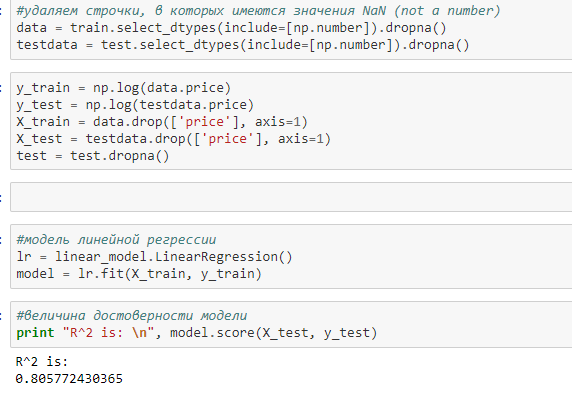
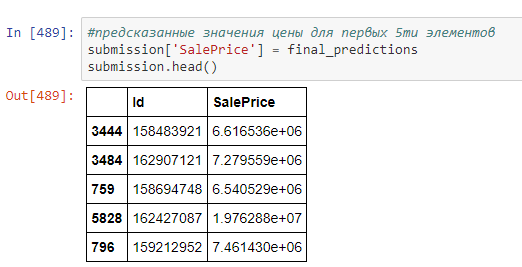


Рисунок 8. Результаты тестирования модели



**Список литературы:**

1. Экономико-математический словарь Изд.5 (2003) , 255с.
2. У. Маккинли. Python и анализ данных, 2015.
3. А.Л.Богданов.URL: «Эконометрический анализ рынка подержанных автомобилей», <http://cyberleninka.ru/article/n/ekonometricheskiy-analiz-rynka-poderzhannyh-avtomobiley>.
4. И.Х. Утакаева, К.А. Хмелевская. Опыт эконометрического анализа Python, Международный научный журнал. Москва**. -**2017- №5.