

УДК 657.1:004

**СОЗДАНИЕ АНАЛИТИЧЕСКИХ ОТЧЁТОВ ПО РЫНОЧНОЙ
СТОИМОСТИ АВТОМОБИЛЕЙ С ИСПОЛЬЗОВАНИЕМ ФУНКЦИЙ
PYTHON**

Ильичев В.Ю.,

к.т.н., доцент,

Калужский филиал ФГОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»,

Калуга, Россия

Сафронова М.Е.,

магистрант,

Калужский филиал ФГОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»,

Калуга, Россия

Аннотация

В статье описано исследование, целью которого являлась разработка программного кода для обработки компьютерных баз данных по статистическим рыночным характеристикам автомобилей различных фирм (в качестве примера рассмотрены Volkswagen и Audi) для получения взаимозависимости выбранных показателей с помощью линейной модели Пирсона. Программа составлена на языке Python, с применением специальных библиотек функций: Pandas, предназначенной для работы с базами данных, Statsmodels для создания и анализа широкого ряда статистических моделей и Matplotlib для вывода полученной в ходе расчётов графической информации.

На примерах бесплатных баз данных по авторынку продемонстрированы все этапы работы программы; по результатам каждого этапа и при окончательном анализе сделаны соответствующие выводы.

Приведено заключение о преимуществах описанного подхода к обработке экономических данных по сравнению с традиционным применением коммерческих программных комплексов и обработке данных «вручную». Показаны достоинства современного универсального языка программирования Python именно в отношении достигнутой цели на примерах получения моделей ценообразования автомобилей фирм Volkswagen и Audi.

Даны рекомендации по развитию разработанной методики.

Ключевые слова: экономические показатели, автомобилестроение, база данных, программа, язык Python, библиотека Pandas, библиотека Statsmodels.

CREATING ANALYTICAL REPORTS ON THE MARKET VALUE OF CARS USING PYTHON FUNCTIONS

Ilichev V.Y.,

PhD, Associate Professor,

Kaluga Branch of Bauman Moscow State Technical University,

Kaluga, Russia

Safronova M.E.,

undergraduate,

Kaluga Branch of Bauman Moscow State Technical University,

Kaluga, Russia

Annotation

The article describes study aimed at developing software code for processing computer databases on statistical market characteristics of cars of various companies

(Volkswagen and Audi as an example) to obtain interdependence of selected indicators using Pearson's linear model. Program was compiled in Python language, using special function libraries: Pandas, designed to work with databases, Statsmodels to create and analyze a wide range of statistical models and Matplotlib to output graphic information.

Examples of free databases on car market demonstrate all stages of program; according to results of each stage and in final analysis, appropriate conclusions were drawn.

Conclusion is given on advantages of described approach to economic data processing compared to traditional use of commercial software complexes and "manual" processing. Advantages of modern universal programming language Python are shown precisely in relation to achieved goal using examples of obtaining car pricing models.

Recommendations on the development of methodology are given.

Keywords: economic performance, automotive, database, program, Python language, Pandas library, Statsmodels library.

Введение. Решение большого количества современных экономических задач связано с обработкой и анализом значительных массивов статистических данных [10]. Такие исследования невозможны без привлечения программных средств и новых технологий программирования [11].

В настоящее время имеется огромное количество программ, специально разработанных для решения разного класса статистических задач [8]. Чаще всего такие программные продукты являются сложными для изучения и дорогими по стоимости. Но, к счастью, существует альтернатива, позволяющая создавать программы для решения практически любой проблемы – применение прикладного программирования, и в данном случае лидером по универсальности и простоте освоения является язык программирования Python

[2] с использованием специальных библиотек функций для реализации необходимых алгоритмов [3]. При рассмотрении экономических проблем выделяются такие его особенности, как возможность работы с огромным количеством форматов данных, а также способов визуализации полученной после обработки информации в различном графическом виде.

Безусловно, на данный момент, автомобильная отрасль (включая производство и продажу автомобилей) является ключевой для экономик большинства развитых стран [9] и от её развития зависит их валовой внутренний продукт. Необходимо отметить, что ситуация на автомобильном рынке является крайне динамичной и нестабильной, поэтому от учёных требуется особенно тщательный анализ статистической информации и разработка на его основе новых всё более совершенных математических моделей [6], а также рекомендаций для их успешного использования.

Цель исследования. В описываемой работе целью исследований является получение наглядных графиков, отображающих влияние разных факторов (характеристик рынка бывших в употреблении автомобилей) на их стоимость. Такие графики просто необходимы для организации служб автомобильных корпораций (автодилеров, отделов маркетинга), имеющих необходимость отслеживания практически в реальном времени складывающейся рыночной ситуации. Например, таким образом, возможно оценить конкурентоспособность одного фирменного знака по сравнению с другим, или одной модели автомобиля по сравнению с другой [1]. Также, полученная наглядная информация поможет покупателям автомобилей по созданным графическим моделям объективно оценить и выбрать оптимальный вариант покупки согласно своим предпочтениям.

Задачами, решаемыми в ходе исследований, являются подбор наиболее удачных библиотек статистического анализа (и библиотек, реализующих прочие функции) для применяемого языка программирования Python, а также

разработка кода программы, позволяющей добиться достижения выше обозначенной цели.

Все применяемые в ходе программирования технологии должны быть проиллюстрированы примерами. В данном случае выбрано создание и сравнение статистических рыночных моделей стоимости для автомобилей двух фирм: Audi и Volkswagen без учёта различия по моделям. В внимание принимается влияние на стоимость только некоторых факторов (к примеру, наиболее важных для покупателя или для отдела маркетинга).

По результатам исследований должны быть сделаны выводы и выработаны рекомендации.

Материал и методы исследования. Для отработки методики был использован пример базовой выгрузка базы данных по продаваемым автомобилям, взятый с сайта [7]. Из полученного файла Excel были выбраны только строки данных, описывающие автомобили фирм Audi и Volkswagen (данные по каждой фирме были выделены в отдельные файлы); также были убраны все лишние столбцы, кроме, тех, влияние которых учитывается на составляемые информационно-математические модели: «Текущая цена», «Год выпуска», «Пробег», «Объём двигателя», «Мощность двигателя». Для фирмы Audi в выгрузке присутствуют данные по 570 автомобилям, для Volkswagen – по 1237.

Основной библиотекой языка Python, использованной для импорта и анализа данных, является модуль Pandas [4], обладающий очень широкими возможностями. Документацию по применению этой библиотеки частично можно обнаружить в сети Интернет [13], однако данной информации, благодаря постоянному совершенствованию входящих в неё функций, часто оказывается недостаточно для практического использования в инновационных проектах. К счастью, благодаря широчайшей распространённости языка Python, существует множество сайтов (к которому, например, относится и сайт автора

[14]) и форумов, восполняющих эти пробелы и позволяющих решить ряд неординарных вопросов.

В результате проведённых исследований была разработана программа на языке Python, обрабатывающая созданные файлы данных по автомобилям, состоящая из следующих блоков:

1. подключение необходимых библиотек команд: основной Pандас, модуля визуализации данных Matplotlib.pyplot и специальной мощной библиотеки вывода статистических моделей в разной форме, позволяющей осуществить их всесторонний анализ Statsmodels [15];
2. осуществление импорта файла базы данных Excel по автомобилям определённой фирмы во внутренний формат Pандас;
3. вывод на экран нескольких первых строк базы данных для проверки (к примеру, 10), включая название параметров, находящихся в каждом столбце;
4. для дальнейшего использования функций пакета статистического анализа – переименование названий используемых параметров в их англоязычный вариант – теперь они называются «Price», «Year», «Mileage», «Capacity», «Power»;
5. использование функции Pандас describe(), формирующей и выводящей на экран и в файл Excel статистические показатели, описывающие загруженные данные;
5. применение функции Pандас corr(), выводящей на экран и в файл Excel коэффициенты корреляции между показателями автомобилей (столбцами), вычисленными по критерию Пирсона, позволяющей выявить степень их линейной взаимосвязи;
6. использование специальной функции пакета Statsmodels:
`smf.ols(formula='Price~Year+Mileage+Capacity+Power',data=df).fit()`,
позволяющей построить комплексную линейную многомерную зависимость стоимости автомобиля от 4 показателей, которую затем можно использовать для предсказания стоимости автомобиля данной фирмы. Однако более

наглядными являются графические зависимости, формируемые далее также пакетом Statmodels.

7. для примера в описываемой программе использован только один из способов построения статистических графиков по созданной линейной зависимости `sm.graphics.plot_fit`, отображающий зависимость стоимости автомобиля от выбранной независимой переменной. Он включает доверительные интервалы прогнозирования, а также отображает истинные значения стоимости, полученные из базы данных (это один из самых простых комплексных статистических графиков). Все возможные способы представления проанализированных статистических данных можно изучить с использованием сайта разработчика пакета Statmodels [15].

Для демонстрации работы созданной программы рассмотрим последовательно процесс анализа статистических данных, полученных для автомобилей Volkswagen и Audi, представленных в виде таблиц Excel, структура которых описана выше. С целью уменьшения объема представляемой информации, приведём только первые 10 строк обрабатываемой базы данных по автомобилям Volkswagen (табл. 1). Дальнейшие стадии анализа, вплоть до финальных результатов, приведём тоже только для автомобилей этой фирмы.

Таблица 1 – Первые 10 строк базы данных по автомобилям Volkswagen.

Текущая цена	Год_выпуска	Пробег	Объем двигателя	Мощность двигателя
375000	2005	226000	2	150
489000	2012	50000	1.6	105
500000	2009	150000	1.8	160
500000	2009	150000	1.8	160
690000	2011	156000	1.8	152
970000	2013	109014	2	140
1745000	2016	29543	2	180
940000	2013	60866	2	170
415000	2012	137000	1.6	105
470000	2012	122000	1.2	105

В табл. 2 представлены статистические показатели, вычисленные в п. 5 программы.

Таблица 2 – Статистические зависимости для базы данных по автомобилям Volkswagen.

	Price	Year	Mileage	Capacity	Power
count	1237	1237	1237	1237	1237
mean	664079,3	2010,948	132587,6	1,843897	138,3154
std	397957,1	4,620789	81774,56	0,47739	41,17519
min	205000	1983	7100	1,2	60
25%	439000	2008	75000	1,6	105
50%	570000	2012	118485	1,6	125
75%	707000	2014	171000	2	152
max	3250000	2019	850000	4,2	350

В данной таблице построчно указаны следующие рассчитанные по всем автомобилям Volkswagen статистические показатели: количество автомобилей в обрабатываемой базе, среднее значение обозначенного в столбце показателя, среднеквадратичное отклонение от среднего, минимальное значение, границы каждого квартиля и максимальное значение.

В следующей табл. 3 представлены коэффициенты корреляции между показателями автомобилей (столбцами), вычисленные по критерию Пирсона.

Таблица 3 – Коэффициенты линейной корреляции для показателей автомобилей Volkswagen.

	Price	Year	Mileage	Capacity	Power
Price	1	0,457251	-0,32606	0,311048	0,429486
Year	0,457251	1	-0,69408	-0,31149	-0,09035
Mileage	-0,32606	-0,69408	1	0,345075	0,085602
Capacity	0,311048	-0,31149	0,345075	1	0,787819
Power	0,429486	-0,09035	0,085602	0,787819	1

Эти данные являются очень наглядными, т.к. явно выявляют тесную связь между некоторыми параметрами (на пересечении каких показателей коэффициент имеет большее значение, там и связь больше). Самая тесная связь наблюдается между мощностью и объёмом двигателя [5]. При этом на цену

оказывают сильное влияние и все остальные 4 показателя (наиболее тесной является зависимость стоимости от года выпуска автомобиля). Естественно, таблица является симметричной относительно диагонали, т.к. в столбцах и в строках фигурируют одни и те же показатели.

Результаты. Так как до этого было выявлено наибольшее влияние года выпуска автомобиля на его стоимость, рассмотрим графически представленные с помощью библиотеки Matplotlib результаты вычисления такого влияния (вычисленного с помощью рассмотренных функций пакета Statmodels) для автомобилей Volkswagen (рис. 1) и Audi (рис. 2).

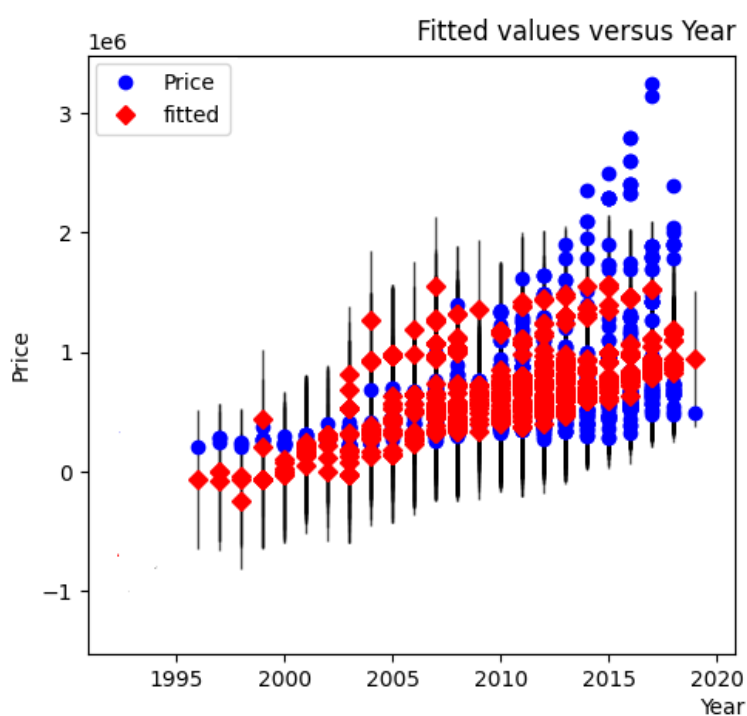


Рис. 1 – Графически представленная зависимость стоимости автомобилей Volkswagen от года выпуска.

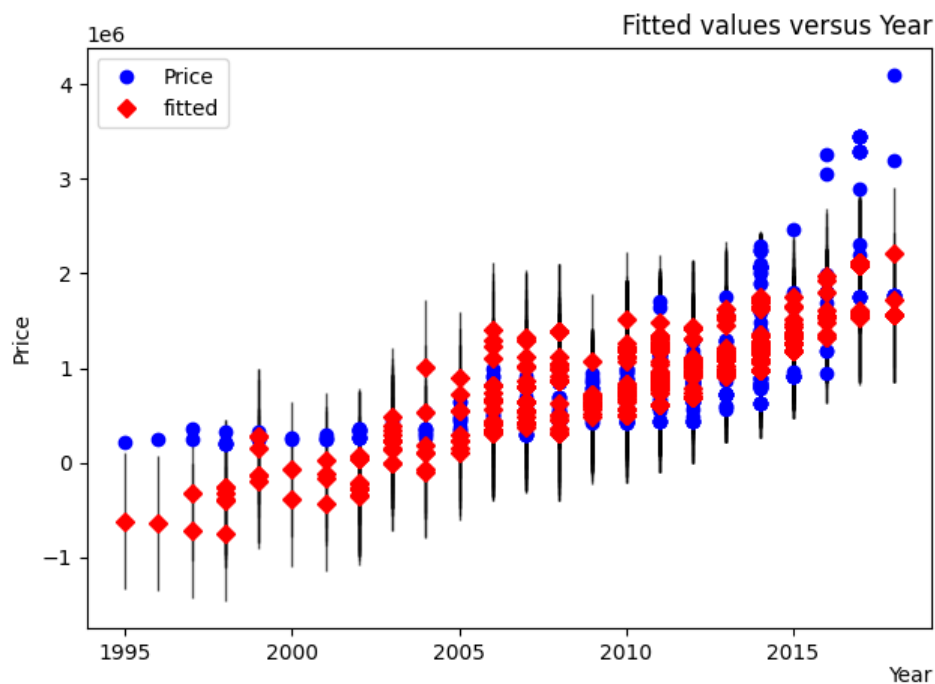


Рис. 2 – Графически представленная зависимость стоимости автомобилей Audi от года выпуска.

На графике синим цветом представлены значения, выведенные из баз данных, а красным – то же количество значений, предсказываемых с помощью созданной программой модели. Так как значения из базы данных отличаются большим разбросом, то и модельные значения также изменяются в довольно широком диапазоне (они ограничены вертикальными линиями – «свечками»). Так как модель является строго математической и линейной, то при большом временном отдалении от текущего момента в прошлое предсказываемые цены могут оказаться даже отрицательными. Конечно, в рассматриваемых случаях такие далёкие промежутки времени не стоит принимать во внимание, а полученные графики следует рассматривать лишь как наглядную иллюстрацию разработанного метода.

Для получения более значимых результатов необходимо, безусловно, использовать более «свежие» и объёмные базы данных. Существуют, например, программы, позволяющие производить парсинг данных с сайта avito, но самой ценной является информация, получаемая непосредственно от дилеров

автоконцернов, и авторы надеются, что их заинтересует представленная в статье технология обработки данных.

Статистические графики, оформленные подобным или иным образом можно использовать при подготовке докладов и презентаций, для размещения в сети интернет, для дальнейшего экономического анализа (например, выявления тренда и прочих закономерностей). Помимо представленных в статье, в использованных в работе библиотеках функций предусмотрены очень широкие возможности создания любого графического оформления результатов исследований.

С целью доработки созданной программы можно оформить для неё пользовательский интерфейс (GUI), реализовать прочие (нелинейные) модели анализа статистических данных и многое другое.

Обсуждение. В статье рассмотрено исследование, использующее самые современные средства создания прикладных программ для статистического анализа. Оно может помочь автостроительным предприятиям и автодилерским центрам сделать для себя вывод, что для них на данный момент открыты возможности получения дополнительной прибыли уже на базе современных научных разработок, а не путём устаревшей на данный момент практически ручной обработки поступающих статистических данных. Следует рекомендовать рассмотренную методику также и «продвинутым» автовладельцам, желающим приобрести автомобиль по наиболее приемлемой цене с учётом всех их пожеланий.

Кроме того, следует иметь в виду, что представленная работа является лишь одним из примеров применения современного языка Python и его библиотек для исследований в области экономики. Данная область традиционно нуждается в создании, хранении и обработке огромных массивов данных, выявление связи между которыми позволяет выявить новые источники прибыли и технического совершенствования продукции. При этом развитие именно рассмотренной отрасли автомобиле- и двигателестроения связано с

непрерывным техническим и социальным прогрессом общества и экономик стран мира, так как такое развитие неразрывно связано с увеличением объёма перемещаемых грузов.

Следует отметить и ещё один крайне важный для менеджеров аспект - для быстрой подготовки наглядных отчётов, отличающихся современным уровнем представления информации, просто не обойтись без программирования с целью автоматизации своего труда. В том числе и этим и объясняется бурное развитие языка программирования широкого профиля Python, характеризующегося простотой освоения и удобством использования для разработки прикладных программ любого назначения.

Заключение. По результатам проведённого исследования можно сделать следующие выводы:

- представленные методы загрузки и обработки обрабатываемых баз данных (практически любого формата), отличаются простотой и широтой выбора дополнительных опций;
- в разработанной программе на языке Python используются всего три программных библиотеки, однако каждая из них имеет огромные возможности (полное описание которых не вошло в статью, но их можно изучить по ссылкам);
- показана простота создания линейных корреляционных зависимостей, доходчиво объяснено значение и возможности применения каждого этапа выполнения написанной программы;
- ввод и вывод результатов осуществляется в виде очень распространённых файлов Excel; выводимые графические результаты являются так же очень наглядными (пользователю рекомендуется попробовать осуществить вывод проанализированных статистических данных и в виде более сложных графиков, показывающих тесную связь между исследуемыми параметрами);
- команды языка Python отличаются простотой освоения и использования.

Рассмотренные в данной статье, а также многочисленные прочие достоинства рассмотренного подхода к анализу статистических данных (в том числе и в автомобильной отрасли) позволяют рекомендовать его к использованию в различных современных отраслях экономики: управлении предприятиями, исследовании закономерностей экономического развития, моделировании пространственной структуры развития экономики и прочих [12]. Авторами данной статьи планируется использование языка Python с библиотеками, написанными на этом же языке, для формирования моделей экономического развития предприятий крупнейшего инновационного региона - Калужской области, отличающейся обилием машиностроительных, и в частности, автомобилестроительных фирм.

Библиографический список:

1. Слободняк И. А., Пискунов И. В. Актуальные проблемы автоматизации бухгалтерского учета и отчетности / И. А. Слободняк, И. В. Пискунов // Бухгалтерский учет в бюджетных и некоммерческих организациях. – 2014. – № 7. – С. 29-34.
2. Сорокина Л. Н. Проблемы внедрения автоматизации учета и подготовки отчетности в условиях перехода на международную систему финансовой отчетности / Л. Н. Сорокина // Финансовая аналитика: проблемы и решения. – 2014. – № 3. – С. 13- 17.
3. Сидорова М. И. Современные информационные технологии как инструмент автоматизации бухгалтерского учета и отчетности / М. И. Сидорова // Международный бухгалтерский учет. – 2011. – № 28. – С. 19-24.
4. Ильичев В.Ю. Разработка программных продуктов с использованием модуля python coolprog для исследования эффективности утилизации тепла продуктов сгорания газообразных топлив. // Системный администратор. - 2020. - № 11 (216). - С. 80-83.

5. Ильичев В.Ю. Использование алгоритма дифференциальной эволюции для решения оптимизационных задач. // Системный администратор. - 2021. - № 4 (221). - С. 80-83.
6. Симушкин А.В., Городничев С.С., Хвалин А.С. Проблемы и задачи автомобилестроения. / В сборнике: Поколение будущего: Взгляд молодых ученых-2019. - 2019. - С. 214-215.
7. Невелев В.А. Применение системного моделирования в прогнозировании конъюнктуры российских рынков автомобильных транспортных средств. / В сборнике: Системное моделирование социально-экономических процессов. труды 41-ой Международной научной школы-семинара. – Воронеж. - 2018. - С. 278-281.
8. Вдовин В.Н. Автоматизация процесса выбора автомобилей рациональной грузоподъемности в логистической системе. // Вестник Восточно-Казахстанского государственного технического университета им. Д. Серикбаева. - 2018. - № 4. - С. 86-94.
9. Парсинг всех полей объявлений с сайтов: Авито, Авто.ру, Юла, Дром. [Электронный ресурс]. - Режим доступа — URL: <https://auto-parser.ru> (дата обращения: 05.11.2021).
10. Ильичев В.Ю., Юрик Е.А. Создание отчетов по доходам организаций с помощью языка Python. // Вектор экономики. - 2020. - № 6 (48). - С. 10.
11. Pandas documentation. [Электронный ресурс]. - Режим доступа — URL: <https://pandas.pydata.org/docs/index.html> (дата обращения: 05.11.2021).
12. Python и энергетика. [Электронный ресурс]. - Режим доступа — URL: <http://turbopython.ru> (дата обращения: 05.11.2021).
13. Statsmodels – introduction. [Электронный ресурс]. - Режим доступа — URL: <https://www.statsmodels.org/stable/index.html> (дата обращения: 05.11.2021).
14. Михайлов Н.А., Тарасюк А.В., Хабардин В.Н., Хабардин С.В. Способ определения эффективной мощности двигателя внутреннего сгорания в

процессе разгона и методика его экспериментальной проверки. // Актуальные вопросы аграрной науки. - 2016. - № 18. - С. 48-54.

15. Шитова Т. Ф. Использование передовых информационных технологий в бухгалтерском учете / Т. Ф. Шитова // Международный бухгалтерский учет. – 2012. – № 22. – С. 21-26.

Оригинальность 84%