**Анализ доступных данных по грузообороту, пассажира обороту и корреляции с долей розничных продаж через интернет в общем объеме оборота розничной торговли**

Для того чтобы сделать предположение влияет ли увеличение доли продаж через интернет на грузооборот автомобильного транспорта, мы можем обратиться к информации, опубликованной на портале открытых данных

<https://data.gov.ru/>

*Портал открытых данных Российской Федерации – это один из ключевых инструментов реализации государственной политики в области открытых данных, которому отводится роль системообразующего элемента, ядра экосистемы открытых данных Российской Федерации.*

<https://data.gov.ru/opendata/7708234640-fiveazeroatwoathreeasix>

Наименование набора данных - Доля продаж через Интернет в общем объеме оборота розничной торговли

*Отношение стоимости товаров, проданных через Интернет, к общему объему оборота розничной торговли. Интернет торговля (электронная торговля) – это продажа товаров по заказам от покупателей, поступивших в интерактивном режиме (ON-LINE), где цена и (или) условия продажи приняты или обговорены по Интернету, электронной почте и т.п., независимо от формы расчета с покупателем и способа доставки товара. Продажа товаров через сеть Интернет, как правило, осуществляется через интернет-магазин. Интернет-магазин - имеющий уникальное доменное имя сайт, позволяющий покупателю заказать и приобрести товар, предварительно ознакомившись с существующей информацией о покупке (товарах, ценах, продавце, месте его нахождении, способах и условиях оплаты, доставке товара и др.). Стоимость товаров, проданных через интернет-магазины, определяется по времени выписки счета или доставки покупателю, независимо от времени фактической оплаты товара покупателем.*

Владелец набора данных - Федеральная служба государственной статистики

Гиперссылка (URL) на набор - <http://fedstat.ru/opendata/7708234640-fiveazeroatwoathreeasix/data-2018-02-06-structure-2018-02-06.xml>

А также к данным с портала федеральной службы государственной статистики <https://rosstat.gov.ru>

*Удовлетворение потребностей органов власти и управления, средств массовой информации, населения, научной общественности, коммерческих организаций и предпринимателей, международных организаций в разнообразной, объективной и полной статистической информации – главная задача Федеральной службы государственной статистики.*

*Для ее решения и действует Росстат, в состав которого входят центральный аппарат на федеральном уровне и территориальные органы, расположенные во всех субъектах Российской Федерации.*

*Международная экспертиза признала статистические данные Федеральной службы государственной статистики надежными.*

*Росстат является учредителем журнала «Вопросы статистики» (Свидетельство о регистрации средства массовой информации номер 012312 от 15 апреля 1994 года выдано Комитетом Российской Федерации по печати Государственному комитету Российской Федерации по статистике (Госкомстат России))*

Данные по грузообороту автомобильного транспорта, по субъектам Российской Федерации.

Данные по грузообороту автомобильного транспорта на коммерческой основе, по субъектам Российской Федерации.

<https://rosstat.gov.ru/statistics/transport>

<https://rosstat.gov.ru/storage/mediabank/Avto-gruz_2021.xls>

*Грузооборот транспорта – объем работы транспорта по перевозке грузов. Определяется суммированием произведений массы перевезенных грузов в тоннах на расстояние перевозки с последующим суммированием указанных произведений по всем позициям перевозки. Расстояние перевозки отдельной партии груза измеряется в километрах, а на морском транспорте в милях (1 миля = 1,852 км). Единицей измерения является тонно-километр (на морском транспорте – тонно-миля).*

*Грузооборот транспорта группируется по видам транспорта, сообщения, ширине колеи и другим признакам. Грузооборот транспорта включает данные как по организациям, для которых эта деятельность является основной (по предприятиям транспорта), так и по организациям других видов деятельности.*

Данные по пассажирообороту автобусов общего пользования по субъектам Российской Федерации.

<https://rosstat.gov.ru/storage/mediabank/t1-7.xls>

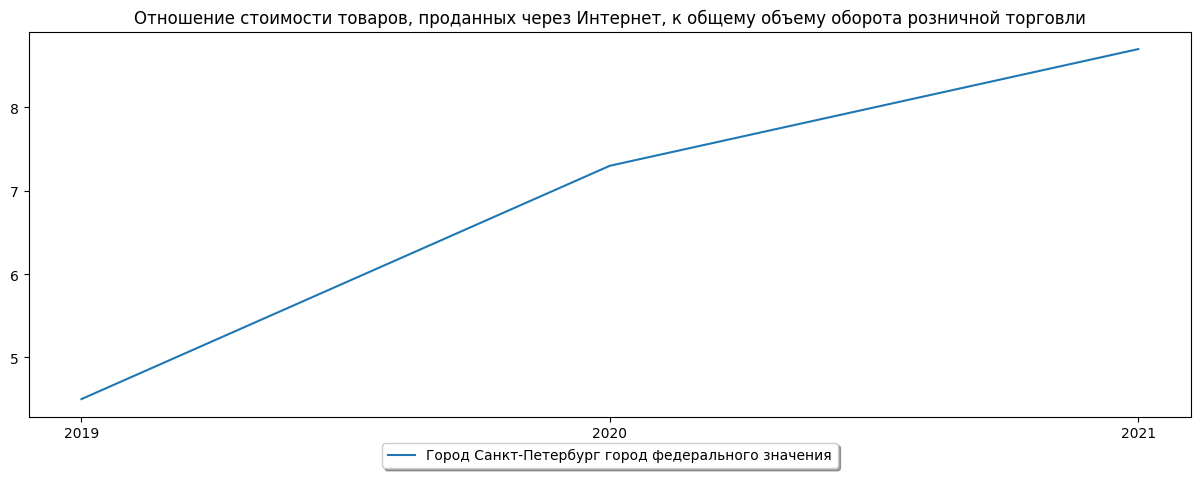
*Пассажирооборот транспорта общего пользования – объем работы транспорта по перевозке пассажиров. Определяется умножением числа пассажиров на расстояние перевозки по каждой позиции перевозки с последующим суммированием указанных произведений. Расстояние перевозки пассажира измеряется в километрах, а на морском транспорте в милях (1 миля = 1,852 км). Единицей измерения является пассажиро-километр (на морском транспорте – пассажиро-миля). Учитывается по видам транспорта и сообщения. Пассажирооборот транспорта общего пользования включает данные как по организациям, для которых эта деятельность является основной (по предприятиям транспорта), так и по организациям других видов деятельности.*

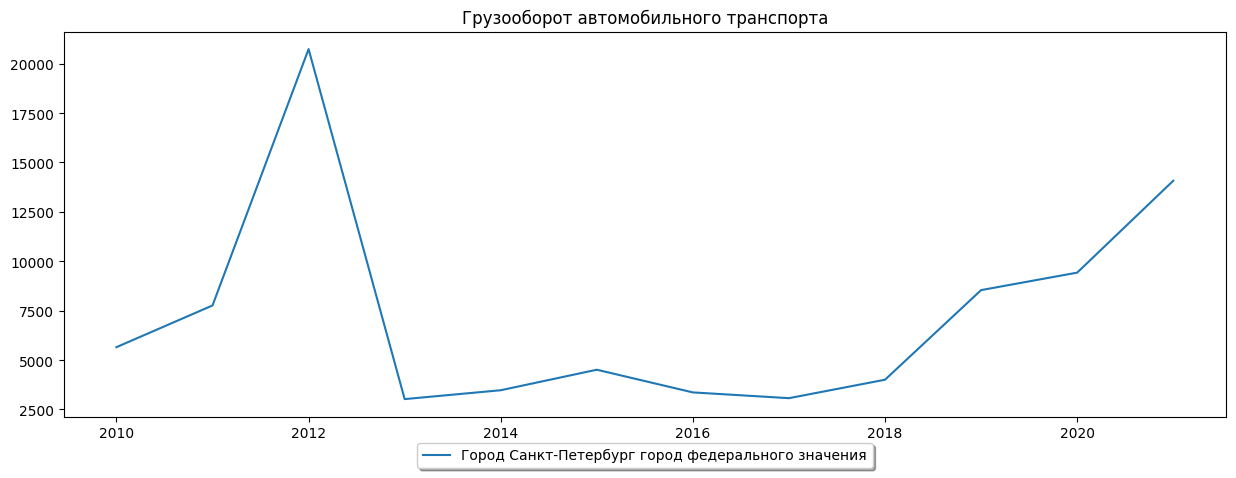
Методологические положения по статистике транспорта - <https://rosstat.gov.ru/storage/mediabank/metod-transp(1)(1)(1).pdf>

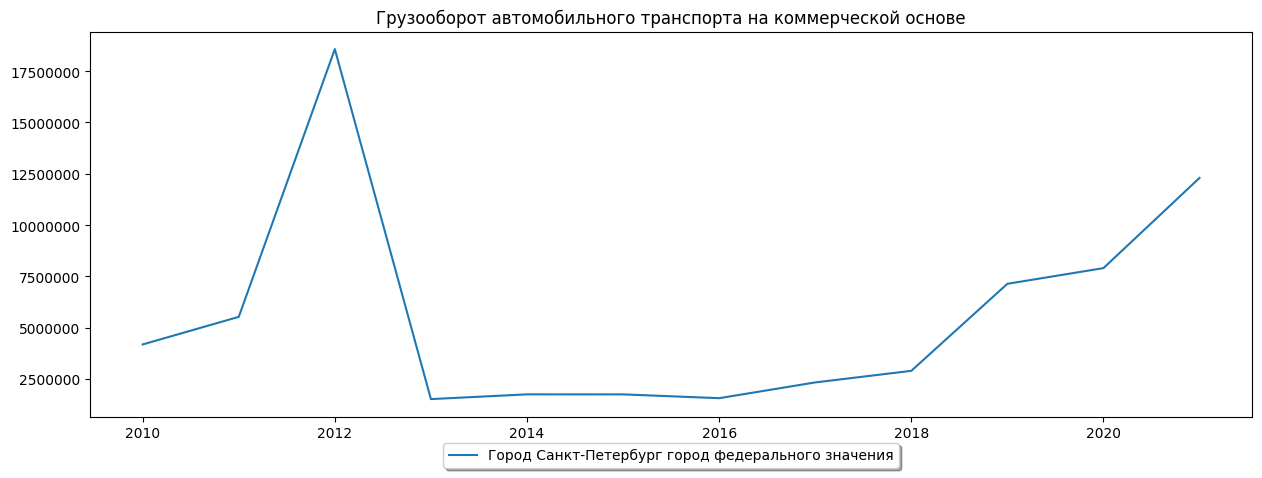
*ГРУЗООБОРОТ ТРАНСПОРТА - объем работы транспорта по перевозкам грузов. Единицей измерения является тонно-километр. Исчисляется суммированием произведений массы перевезенных грузов в тоннах на расстояние перевозки в километрах (милях). Грузооборот транспорта группируется по видам транспорта, сообщения, ширине колеи, роду грузов и другим признакам.*

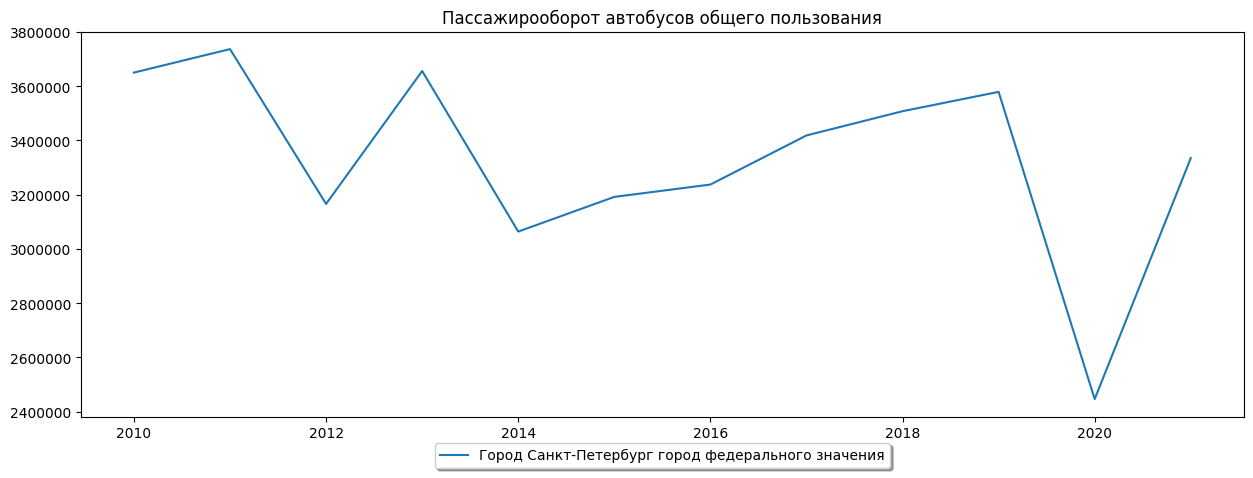
*ПАССАЖИРООБОРОТ ТРАНСПОРТА - объем работы транспорта по перевозкам пассажиров. Единицей измерения является пассажирокилометр, т.е. перемещение пассажира на расстояние в 1 км. Определяется суммированием произведений количества пассажиров по каждой позиции перевозки на расстояние перевозки; вычисляется раздельно по видам транспорта, сообщения и другим признакам. Перевозки грузов, грузооборот, перевозки пассажиров, пассажирооборот транспорта включают данные как по организациям, для которых эта деятельность является основной (по предприятиям транспорта), так и по организациям других видов деятельности.*

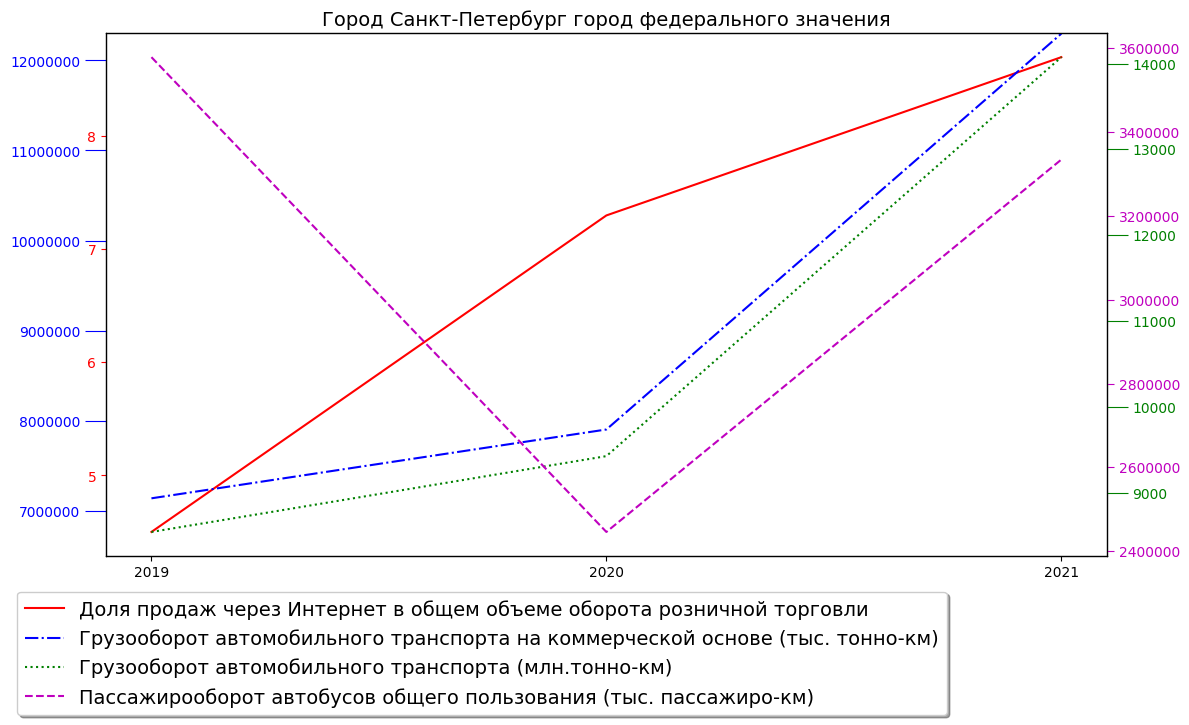
Попробуем сопоставить данные по Санкт-Петербургу за доступный период.

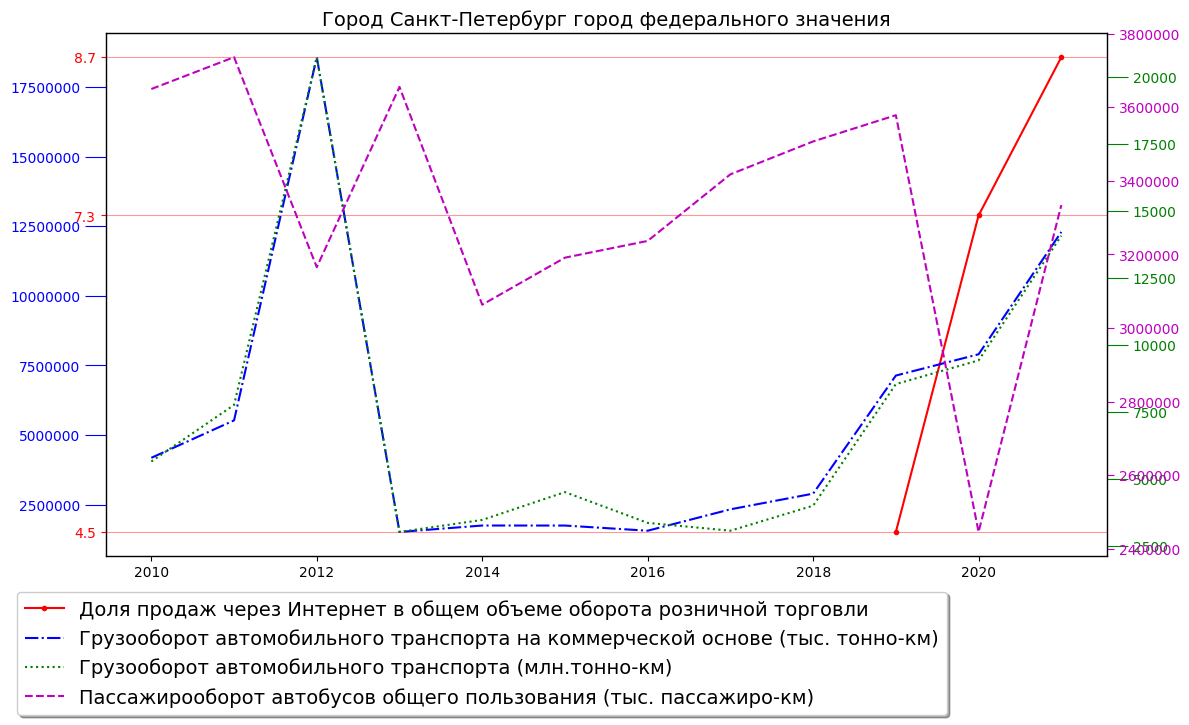










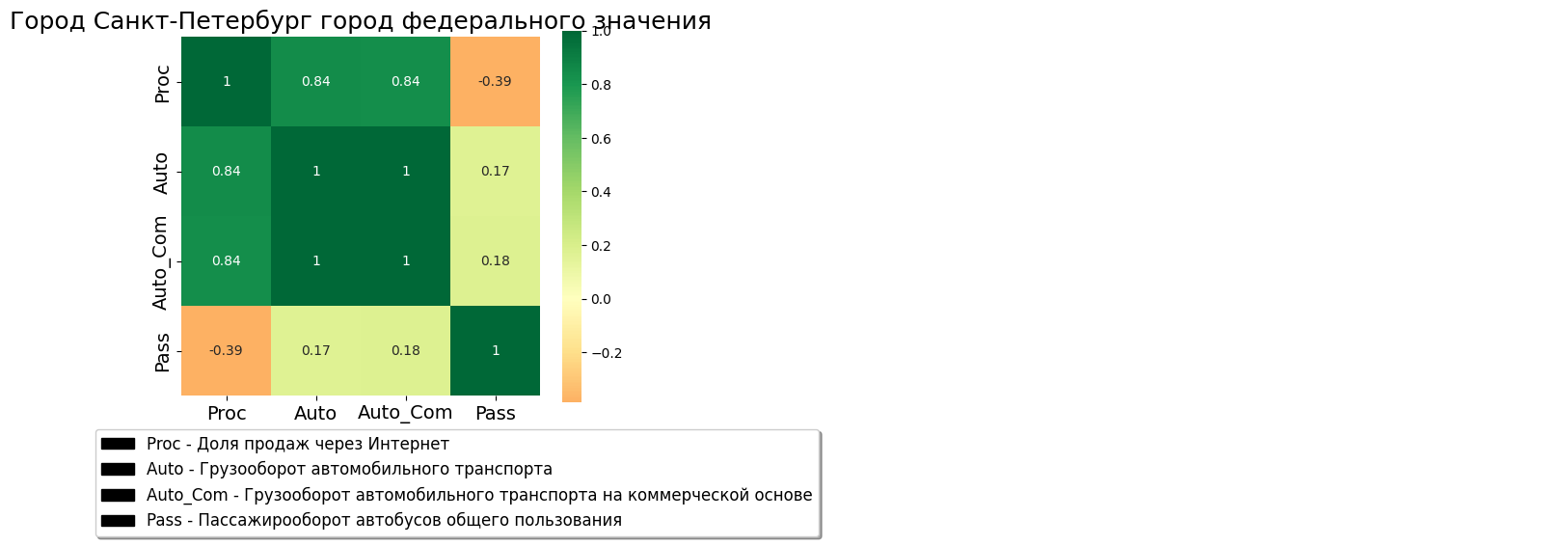


Интуитивно можно предположить, что значения грузооборота и доли продаж через интернет могут быть ковариационно связаны (положительная ковариация) и иметь противоположное значение ковариации между долей продаж через интернет и пассажирооборотом.

Несмотря на то что мы имеем небольшое количество наблюдений для каждого региона попробуем рассчитать корреляции для этих значений используя Линейный коэффициент корреляции r-Пирсона

*Коэффициент корреляции Пирсона (r-Пирсона) применяется для исследования взаимосвязи двух переменных на одной и той же выборке. Он позволяет определить, насколько пропорциональная изменчивость двух переменных.*

*Другими словами, это ковариация двух переменных, деленная на произведение их стандартных отклонений. Форма определения включает в себя "момент произведения", то есть среднее значение (первый момент относительно начала координат) произведения случайных величин с поправкой на среднее значение.*

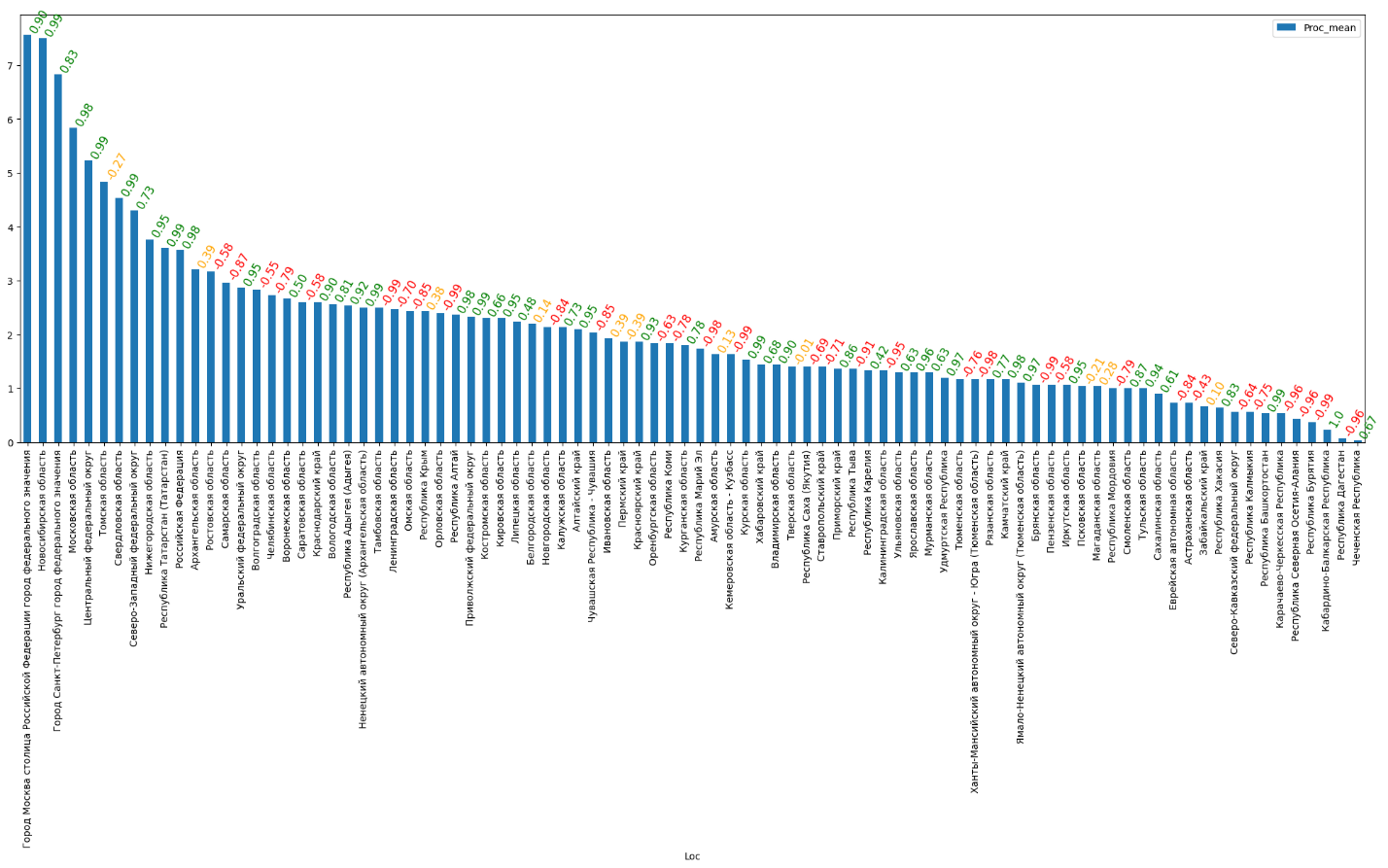


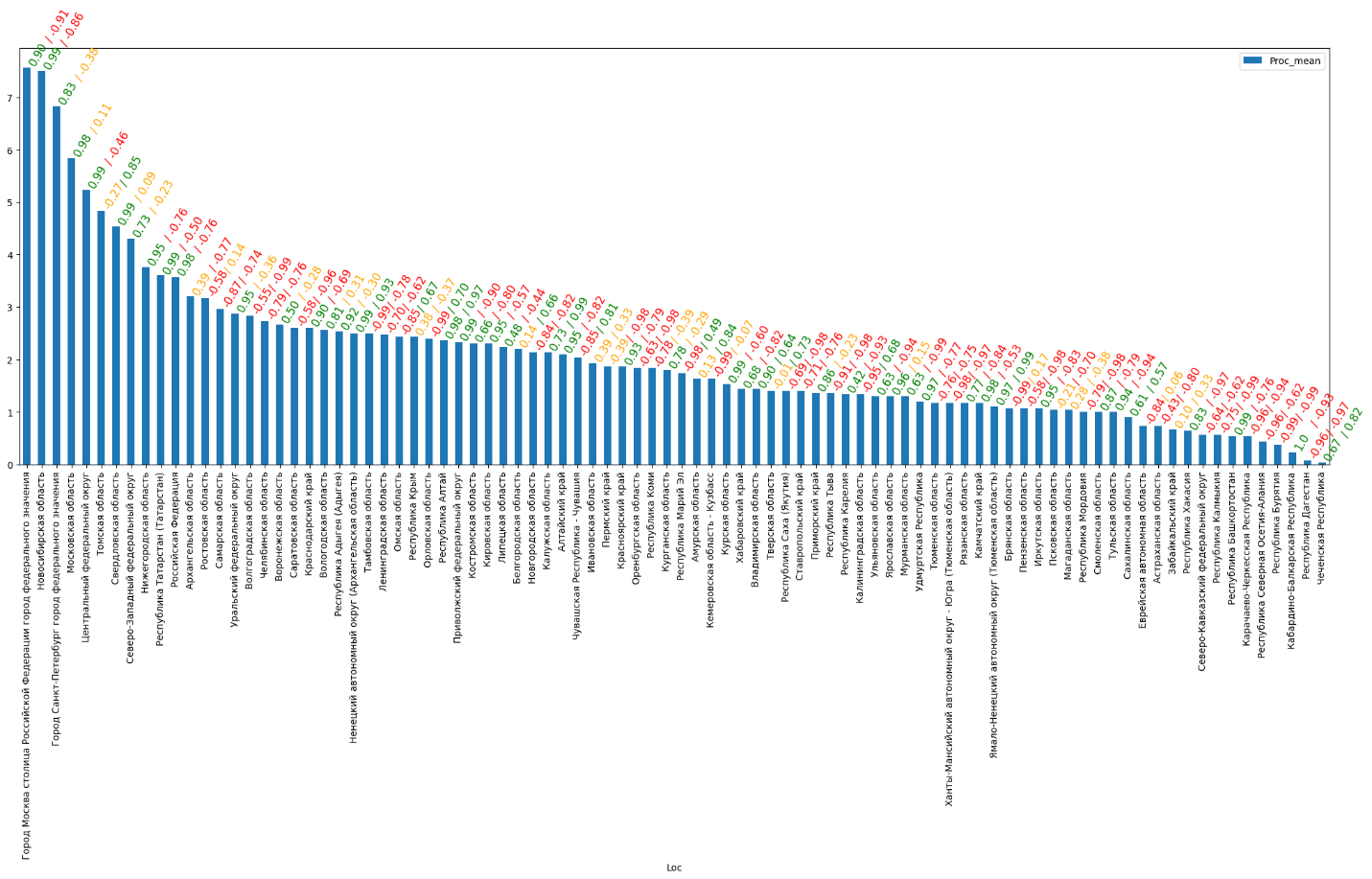
Существуют различные рекомендации по интерпретации коэффициента корреляции, предложенные различными авторами. Однако интерпретация коэффициента корреляции часто зависит от контекста и целей исследования. Так коэффициент корреляции 0,8 может расцениваться как очень низкий если проверять какие-либо физические законы на высококачественных данных, но также может рассматриваться и как очень высокая в тех случаях, когда на наблюдения влияют множество различных факторов. В общем случае чтобы результат попал в доверительный интервал при малом количестве наблюдений коэффициент корреляции должен быть достаточно высоким. Также стоит упомянуть о том, что для двух взаимосвязанных переменных причинно-следственная связь не может быть дана как должное.

В данном случае мы скорее всего имеем большое количество различных сложных факторов, влияющих на грузооборот. К сожалению, данные при суммировании по периодам утрачивают большое количество информации, которая позволяла более точно выявить взаимосвязь. Тем не менее показатели автомобильного грузооборота выглядят взаимосвязанными, однако несмотря на высокую корреляцию, при малом количестве наблюдений мы не можем уверено сказать, что рост процентного отношение стоимости товаров, проданных через Интернет, к общему объему оборота розничной торговли влияет на объёмы перевозок автомобильным транспортом. Также видим хотя и слабую, но всё же отрицательную корреляцию с пассажирооборотом автобусов общего пользования, по всей видимости на этот показатель оказывает сильное влияние ковидные ограничения введённые в этот период, по завершении которого наблюдается восстановления значений пассажирооборота что не позволяет нам увидеть сильную отрицательную корреляцию.

Немного поразмыслив мы можем сделать предположение что чем больше процент стоимости товаров, проданных через Интернет, по отношению к общему объему оборота розничной торговли тем больше будет прослеживаться взаимосвязь между наблюдаемыми переменными, то есть для тех регионов где высокий процент стоимости товаров проданных через Интернет мы ожидаем увидеть более высокую положительную корреляцию с автомобильным грузооборотом и заметную отрицательную корреляцию с пассажирооборотом

Для этого попробуем проанализировать данные по всем доступным регионам отсортировать их по значению доли продаж через Интернет в общем объеме оборота розничной торговли и вычислить значение корреляций для всех регионов.





Построив соответствующий график мы действительно можем наблюдать что в тех регионах где электронная торговля наиболее развита получены высокие показатели положительной корреляции Доля продаж через Интернет в общем объеме оборота розничной торговли с Грузооборот автомобильного транспорта на коммерческой основе и не прослеживается зависимость с пассажирооборотом автобусов общего пользования что по всей видимости связано либо с отсутствием данной взаимосвязи либо с сильным влиянием ковидных ограничений на этот показатель.

И одновременно наблюдается некоторая аномалия для Томской области.

Построим графики чтобы детальнее ознакомиться с данными по г. Москве которая имеет наибольшее значение доли продаж через интернет и с данными по Томской области.

Ознакомившись с данными по Томской области, мы видим здесь некоторые особенности быстрый рост Доля продаж через Интернет в общем объеме оборота розничной торговли с 2019 по 2020 с 3.3 % и никакого роста с 2020 по 2021 год ~~~~ что может свидетельствовать об ошибке в данных либо какой-то аномалии. ~~~~

График же по г. Москва хорошо иллюстрирует наши ожидания взаимосвязи автомобильного грузооборота и Доля продаж через Интернет в общем объеме оборота розничной торговли.

