

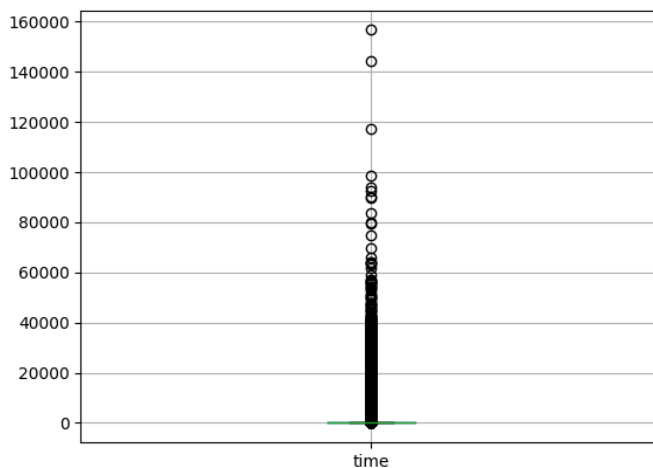
Summary 06.04.2025

Чистка данных

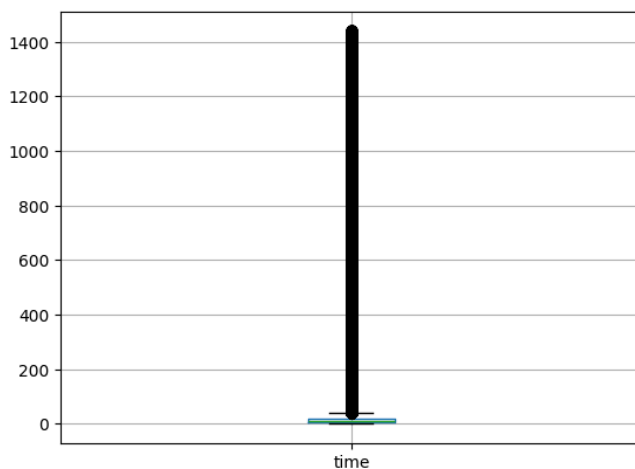
Было обнаружено **1054** строки **явных** дубликатов. Они удалены.

Удалены поездки, которые **начались позже чем закончились** или длились **0** секунд - 14291 строк

Распределение поездок в минутах:



Длинные поездки возможны, но являются **редкостью** (0.097% от всех поездок). Также максимальное время бесплатной поездки по подписке - **45** минут. Следовательно, сервис нацелен больше на **короткие** поездки.



Поэтому удалим поездки большие **1440** минут (1 день).

Теперь откинем **2,5%** с **каждой стороны**, т.е. поездки > 71.83 и < 1.3 минуты.

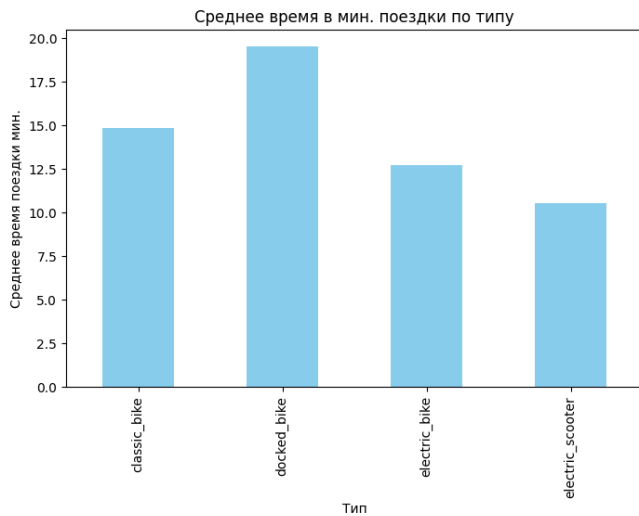
Поездки < 1.3 минуты поместим в **отдельную** группу. Возможно здесь есть поездки, когда

людям **не понравился** велосипед и они его **вернули**.

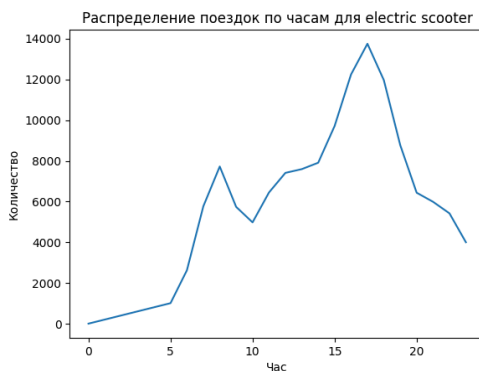
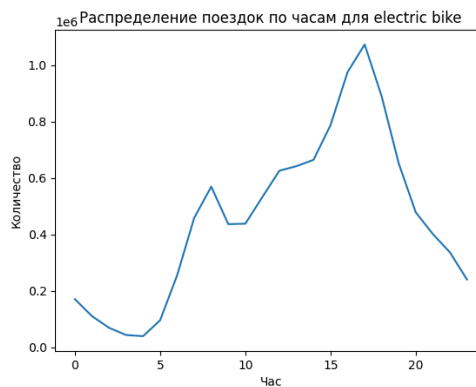
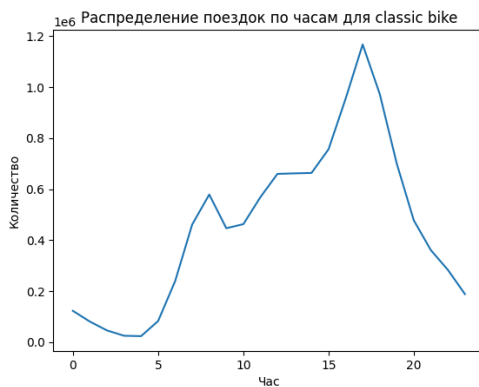
Разведывательный анализ

Рассмотрим на что может влиять тип велосипеда.

Можно сделать вывод, что на **docked bike** совершается в среднем самые **длительные** поездки, при этом на **electric scooter** самые **короткие**.

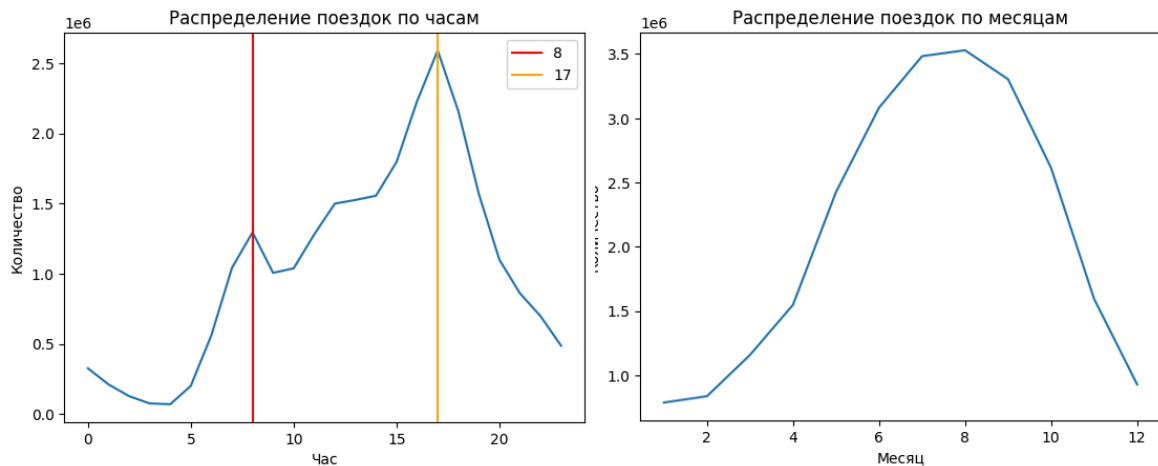


Рассмотрим, есть ли влияние на распределение поездок по часам суток. Можно заметить, что такого влияния **нету**.

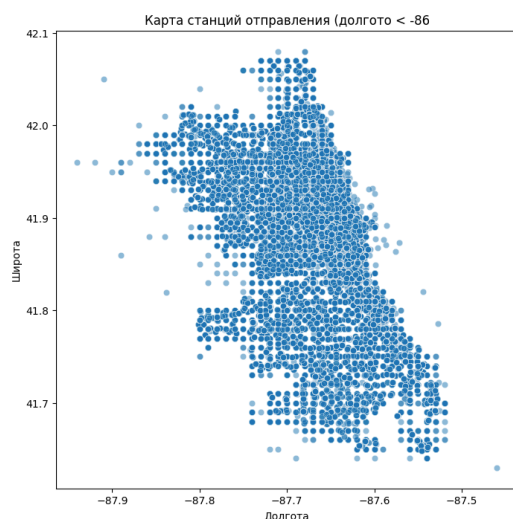


Рассмотрим общую ситуацию, заметим, что основные **пики** приходят на **8** утра и **17** часов вечера, что логично, потому что в это время люди едут на работу и с работы.

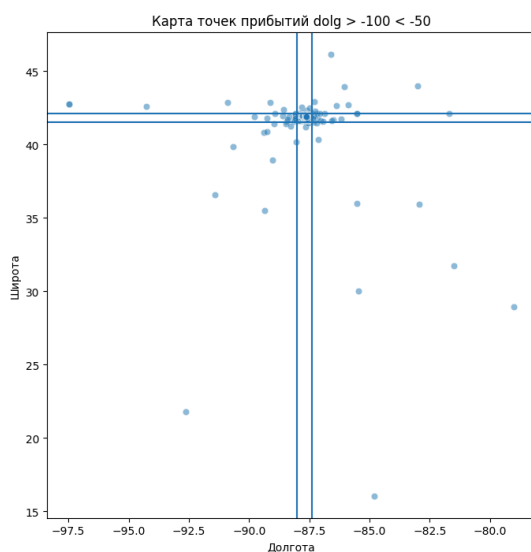
Также рассмотрим, как сезонность влияет на наши поездки. Заметим, что все логично, конец весны, лето и начало осени – **основное** время совершения поездок.



Рассмотрим, где в основном находятся станции.

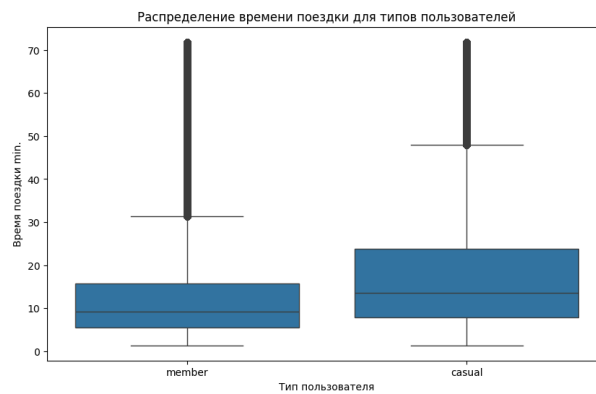


Можно заметить, что **все** стартовые станции находятся **близко, кроме одной**, которая находится рядом с **Монреалем** (1199 км от Чикаго).

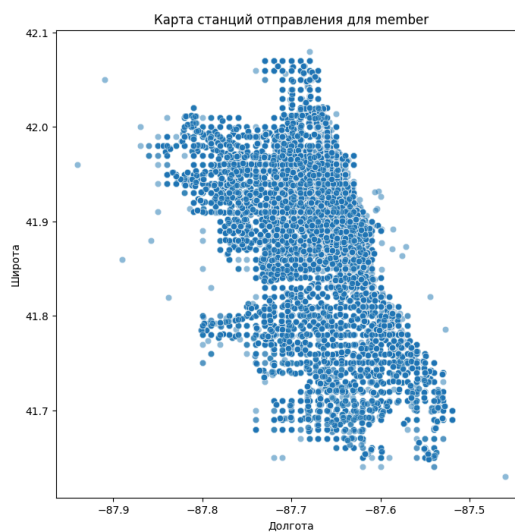
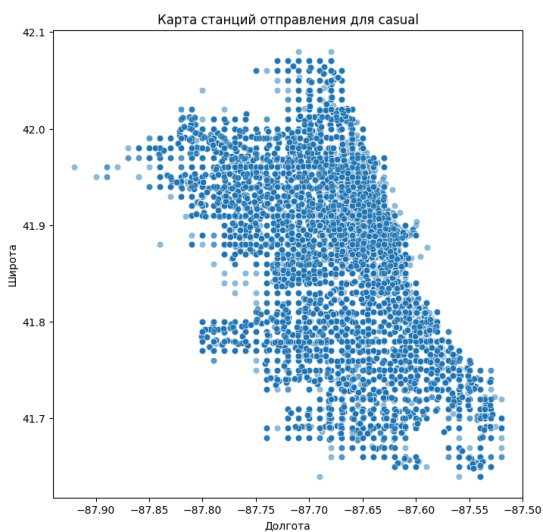


Теперь рассмотрим станции прибытия, заметим, что они находятся **более широко**, чем станции отправления. В прямоугольник, ограниченный четырьмя линиями, попадают все стартовые станции.

Теперь рассмотрим, как тип подписки (member/casual) влияет на поездки. По распределению можно заметить, что поездки людей **без подписки** обычно **длиннее**, чем у подписчиков.



Можно заметить, что у **подписчиков** два **пика** – **8** утра и **17** вечера. Однако, у **casual** **пик** только в **17** вечера. Рассмотрим откуда обычно уезжают люди с разной подпиской.



Можно заметить, что карты особо не различаются, поэтому можно сказать, что **во всех районах есть оба вида пользователей.**

Рассмотрим, есть ли круговые поездки (поездки у которых начальная и конечная точка совпадает). Таких поездок **3.7%** от общего числа. Как идея, можно считать за **неудачную поездку**, поездку которая **< 1.3 минуты и круговая.**