

# Т-Банк: поездки на самокатах



Разведочный анализ данных  
и продуктовые гипотезы для  
оптимизации прибыли и  
улучшения  
пользовательского опыта

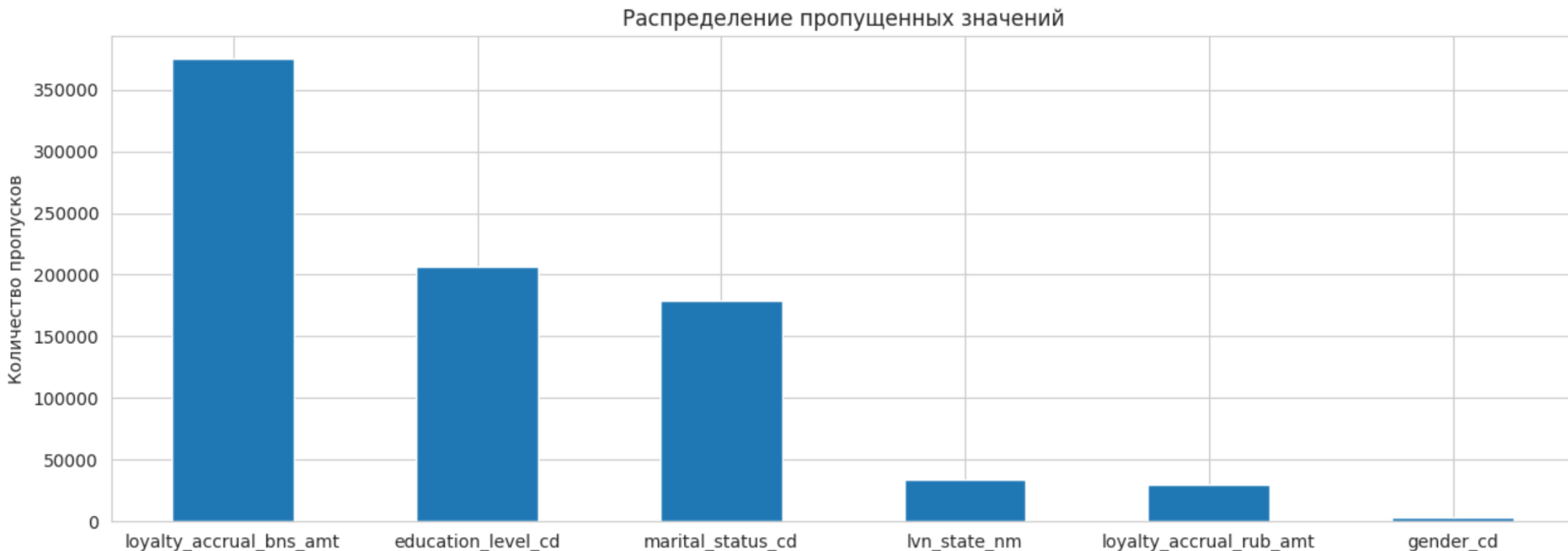
# Разведочный анализ данных (EDA)



## Общая статистика:

В представленном датасете около 400 тысяч наблюдений (объектов) и 20 столбцов с различными признаками. Среди них 4 признака были переведены в формат с датой/временем, 5 категориальных признаков и 11 числовых.

Был рассчитан процент пропущенных значений в каждом столбце. Распределение пропусков представлено на рисунке:

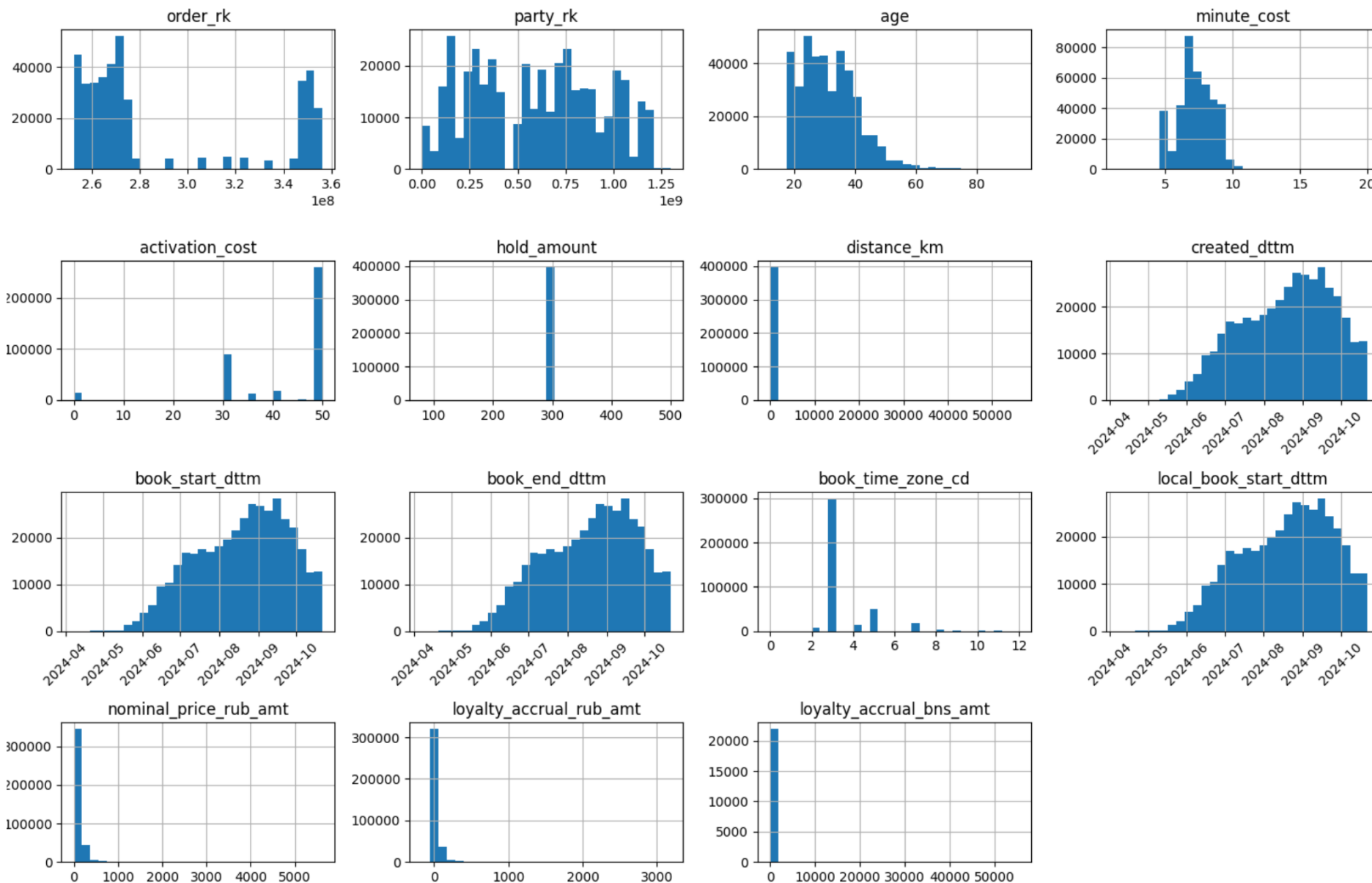


# Разведочный анализ данных (EDA)

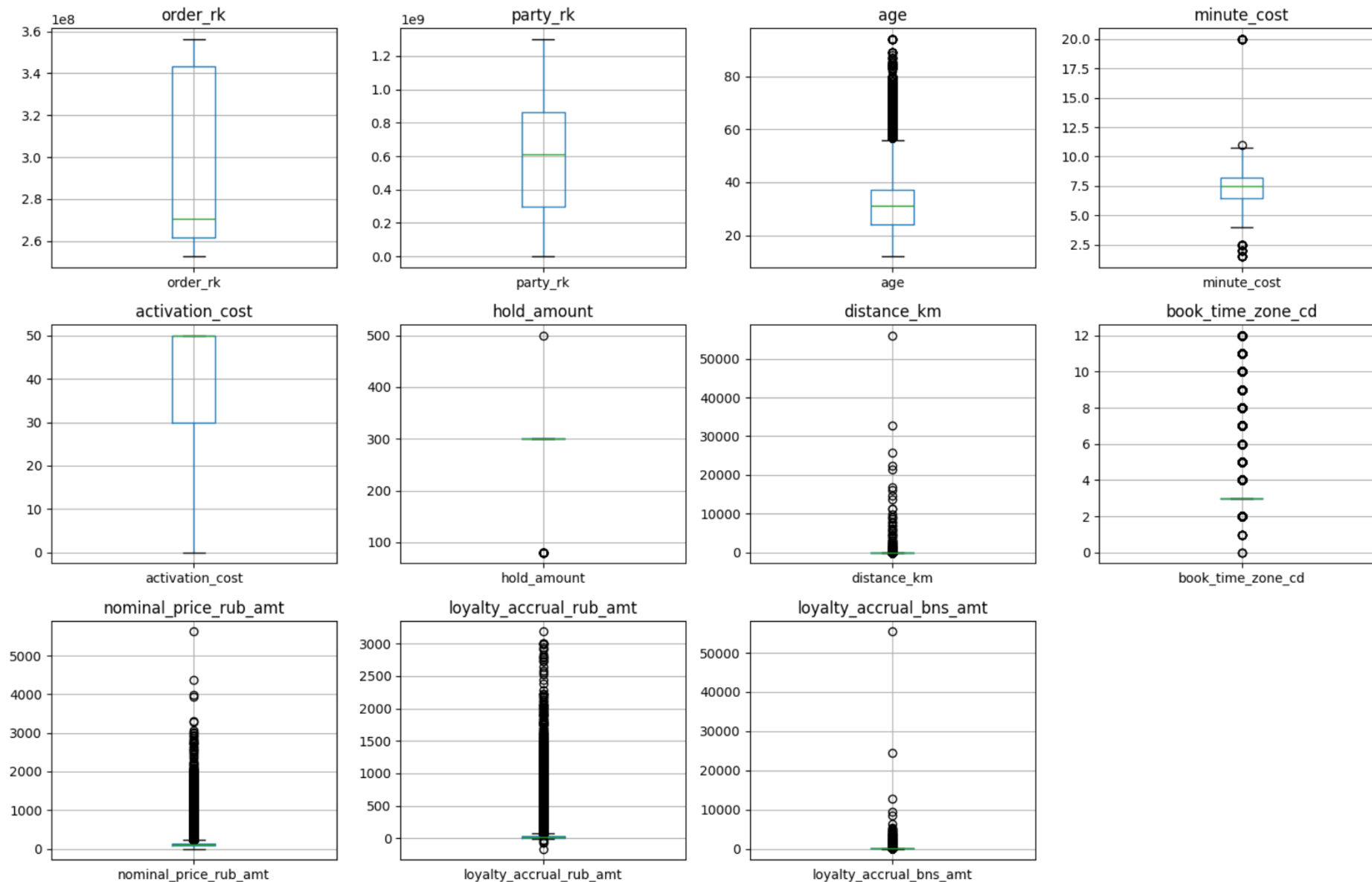


## Анализ числовых признаков:

На рисунке представлены распределения числовых признаков. Наблюдаются различные типы распределения



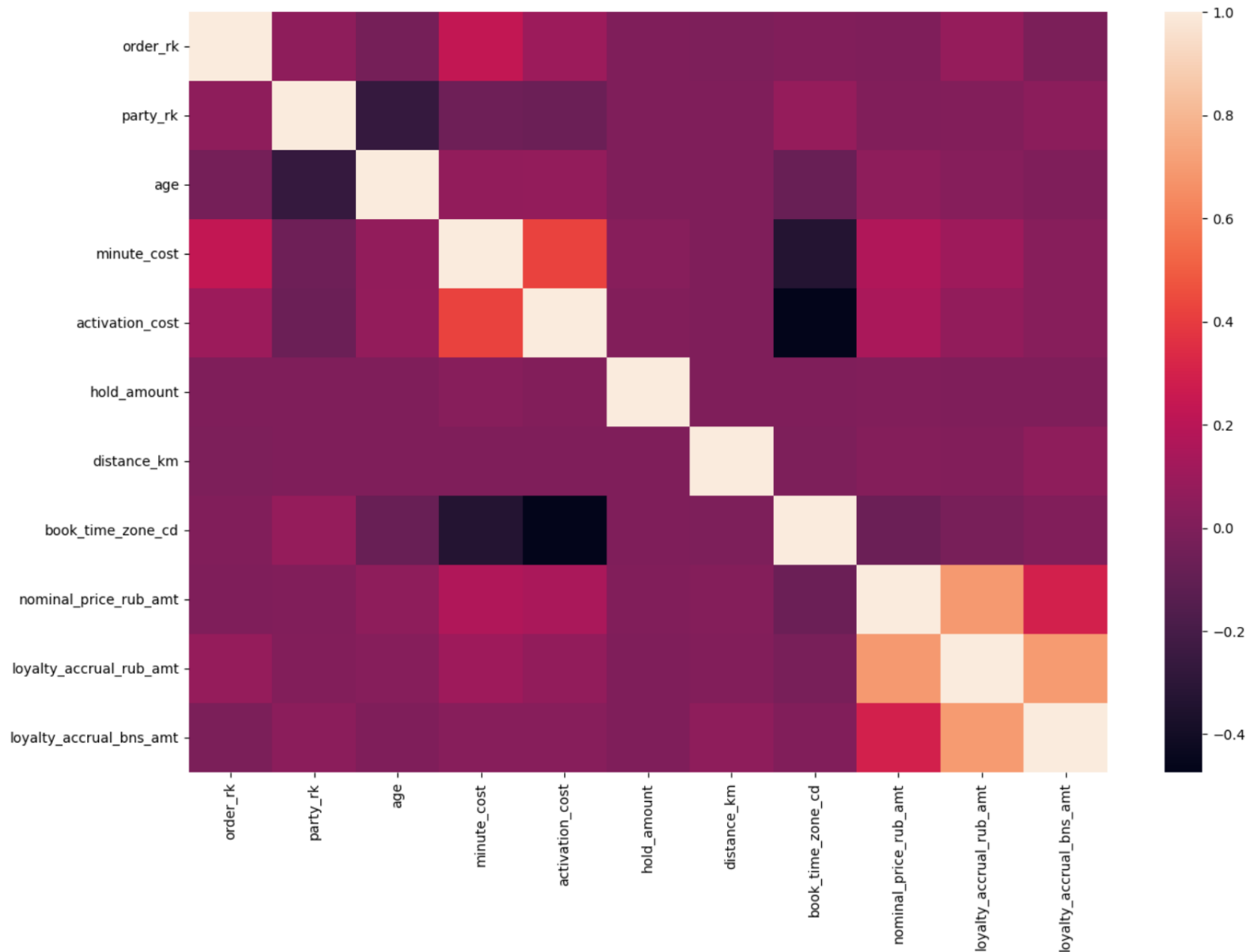
# Разведочный анализ данных (EDA)



## Анализ числовых признаков:

На рисунке представлены ящики с усами для числовых признаков. Видно, что колонки nominal\_price\_rub\_amt, loyalty\_accrual\_rub\_amt, loyalty\_accrual\_bns\_amt, age и distance\_km содержат много выбросов

# Разведочный анализ данных (EDA)



## Анализ числовых признаков:

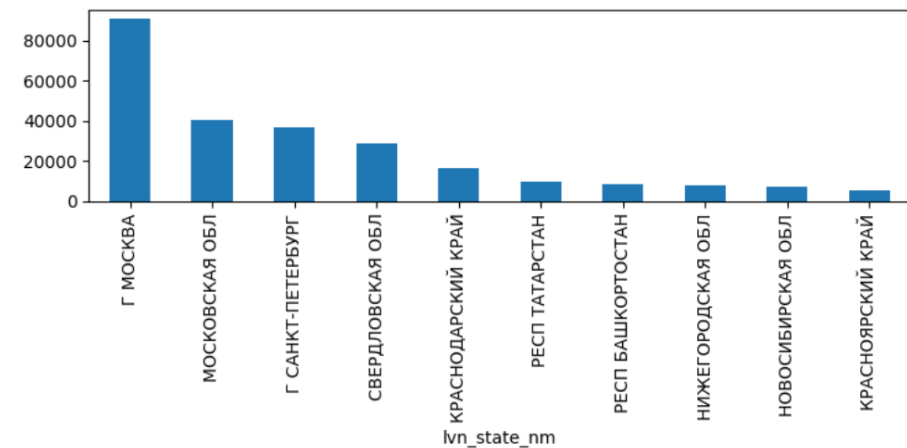
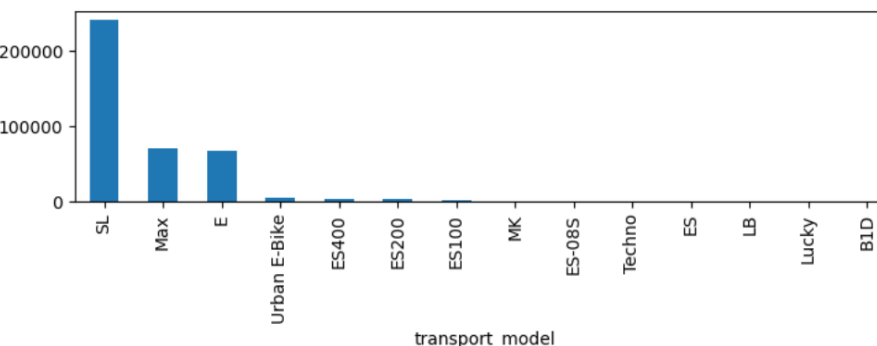
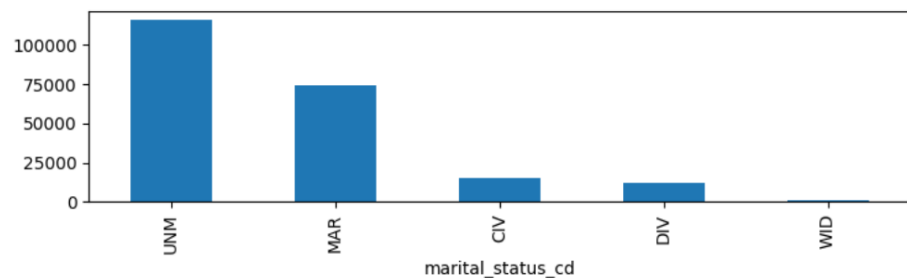
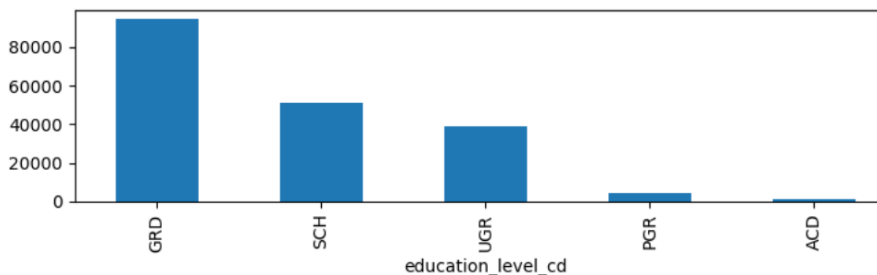
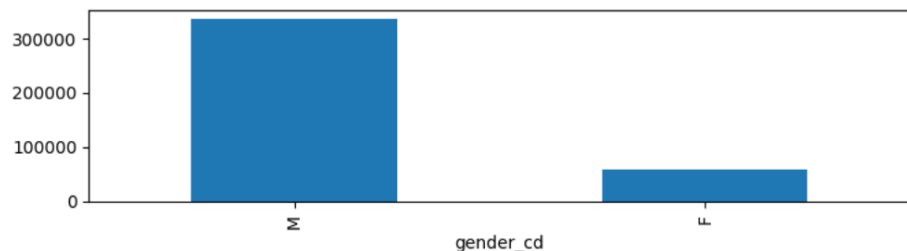
На рисунке представлена тепловая карта для числовых признаков.

Видно, что размер кэшбека в рублях и бонусах хорошо коррелируют между собой и со стоимостью поездки. Это согласуется с действительностью.

# Разведочный анализ данных (EDA)



Распределение категориальных признаков

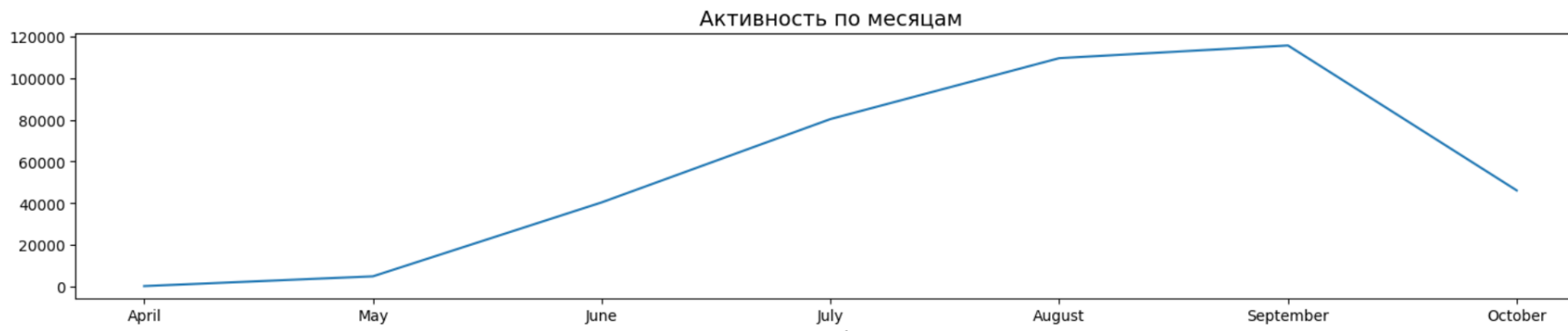
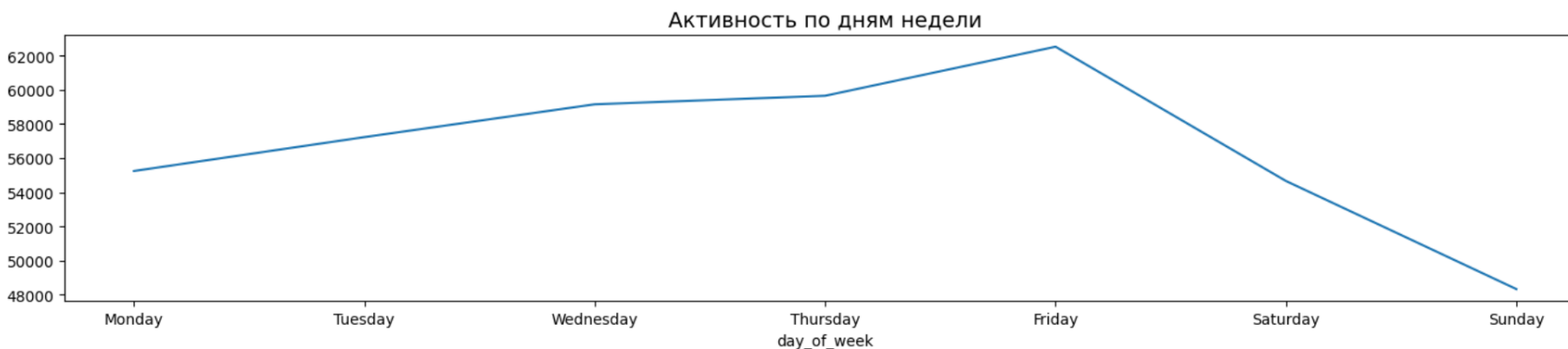
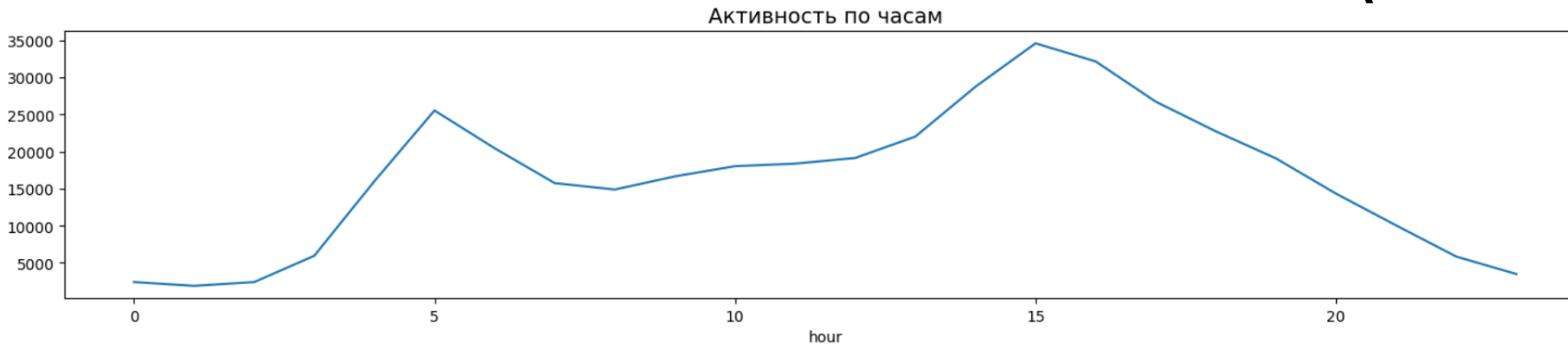


## Анализ категориальных признаков:

На рисунке представлены распределения для категориальных признаков. Видно, что пользователями в основном являются мужчины.

Географическое распределение показывает концентрацию спроса в Москве. Особый интерес представляет анализ семейного статуса: среди пользователей преобладают клиенты, не состоящие в браке.

# Разведочный анализ данных (EDA)



## Сезонность использования сервиса:

По времени суток – пики утром (8-10) и вечером (17-19), ночью спроса почти нет

По дням недели – равномерное распределение, без выраженных различий

По месяцам – высокий спрос летом (май-сентябрь), зимой (ноябрь-март) резкий спад

**Вывод:** сервис востребован круглосуточно в сезон, но зависит от времени суток.

# Продуктовые гипотезы



## *Гипотеза 1.*

Введение скидки 10% для клиентов с высшим образованием (`Education_level = 'GRD'`) увеличит частоту использования сервиса на 15%.

Обоснование: Клиенты с высшим образованием могут быть более лояльны и чаще использовать сервис при наличии выгоды.

## *Гипотеза 2.*

Уменьшение времени ожидания подтверждения аренды с 30 секунд до 10 секунд снизит количество отказов на 20%.

Обоснование: Скорость обработки заявок влияет на удовлетворенность клиентов.

## *Выбор перспективной гипотезы:*

Перспективной является гипотеза 2, так как ускорение обработки заявок напрямую влияет на пользовательский опыт. К тому же, данную гипотезу можно легко проверить с помощью A/B-теста, разделив пользователей на две группы: с текущим временем ожидания и с уменьшенным.



# Доля Т-Банка в индустрии



Для оценки доли Т-Банка в индустрии аренды самокатов можно сравнить количество поездок в датасете с общим рынком. Либо проанализировать среднюю стоимость поездки и сравнить с конкурентами. Это возможно сделать, если будут доступны данные других компаний.