

INOPOLIS  
UNIVERSITY



## Анализ выручки придорожного кафе

innopolis  
UNiversity

innopolis  
UNiversity

innopolis  
UNiversity

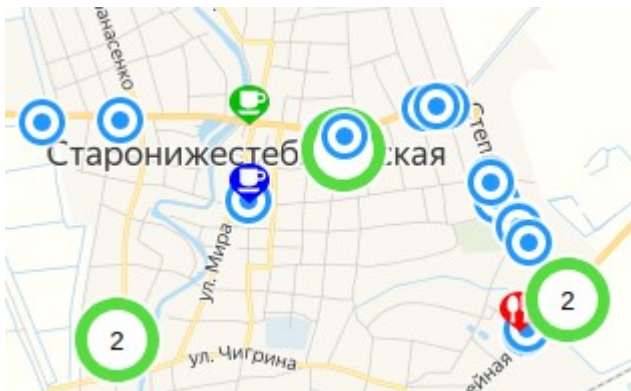
**Цель:** оценить финансовые поступления предприятия в текущем и будущем периоде в условиях пандемии короновирусной инфекции и существующей конкуренции

**Задачи в рамках исследования:**

- провести анализ выручки придорожного кафе;
- выбрать модель прогнозирования;
- построить прогноз выручки на предстоящий период.

# Объект исследования

Выбранное кафе находится на трассе Тимашевск - Славянск-на-Кубани и ведет свою деятельность в условиях жесткой конкуренции. На участке дороги, протяженностью 4 км, находятся 10 придорожных кафе и 10 продуктовых магазинов.



# Предобработка данных

Полученный файл содержит данные о кассовой выручке с 27 февраля 2020г по 2 марта 2022г.

Датафрейм содержит 9 полей:

*date* - дата

*colkas* - количество операций за наличный расчет

*sumkas* - сумма операций за наличный расчет

*colbez* - количество операций по карте

*sumbez* - сумма операций по карте

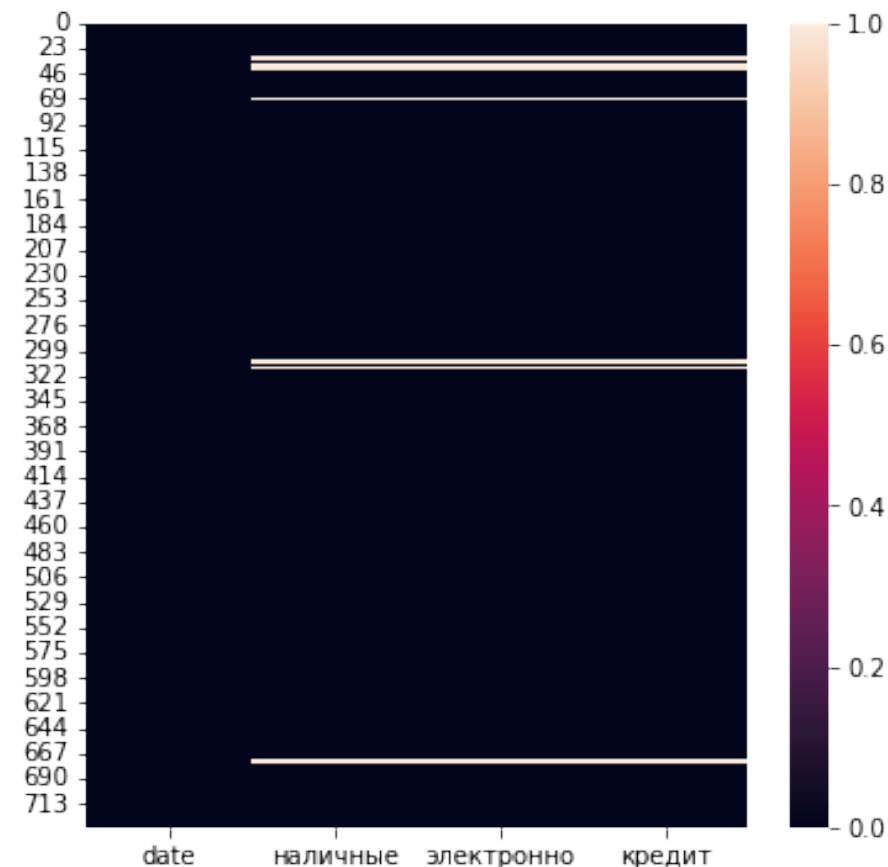
*colskid* - количество предоставленных скидок

*sumskid* - сумма предоставленных скидок

*colkredit* - количество операций в кредит или по договору

*sumkredit* - сумма операций в кредит или по договору

Входной файл содержит 20 пропущенных дней. Временной ряд восстановлен, отсутствующие значения заменены на 0.





# Статистическая информация

Выручка состоит из оплат наличными и электронно картами или через платежные системы GooglePay, ApplePay и т. д. Также присутствуют операции по договорам или в кредит. Поступления неравномерны в течение года и увеличиваются в летний период.



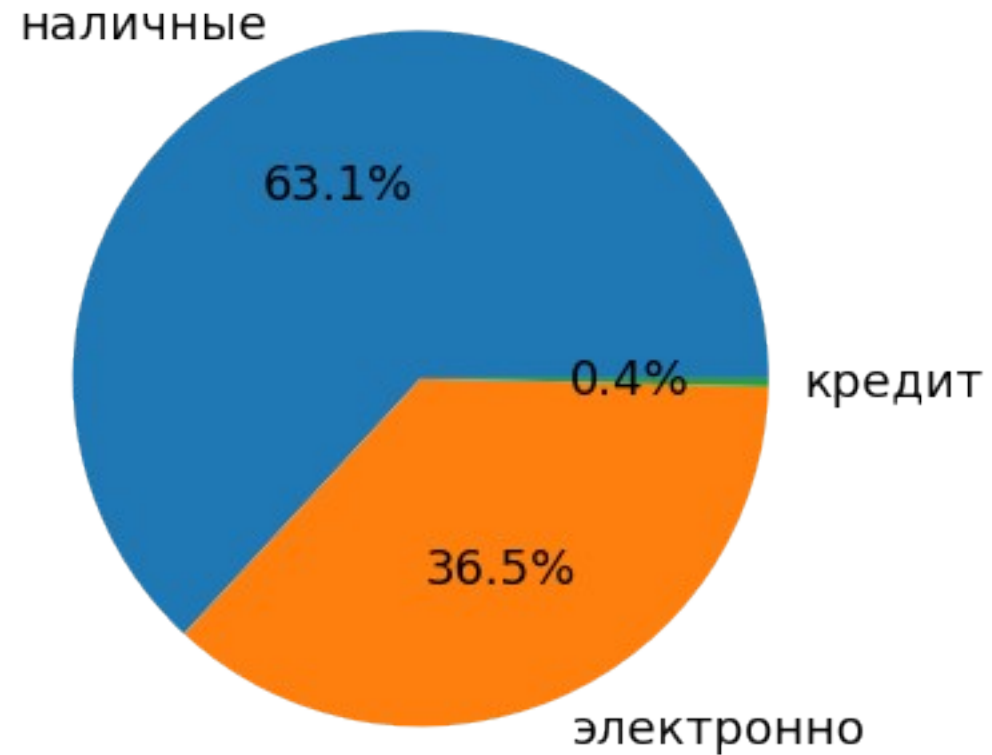
# Статистическая информация

## Соотношение выручки по видам поступлений

Наблюдается дисбаланс в видах оплаты.

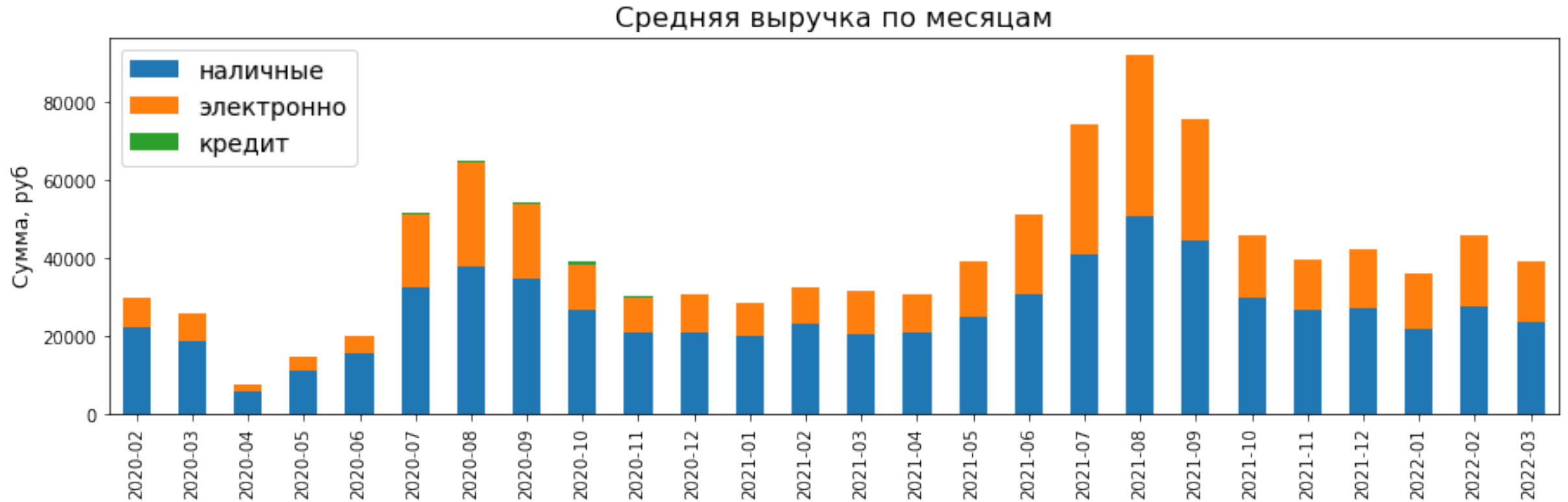
Сумма оплаты наличными составляет приблизительно  $\frac{2}{3}$  от общей суммы выручки.

В условиях пандемии рекомендуются бесконтактные способы оплаты. Но в настоящее время они составляют только  $\frac{1}{3}$  всех поступлений.



# Статистическая информация

Максимальная выручка в течение года приходится на июль, август и сентябрь. В апреле и мае 2020г наблюдается снижение поступлений из-за введенных ограничительных мер в связи с коронавирусом.



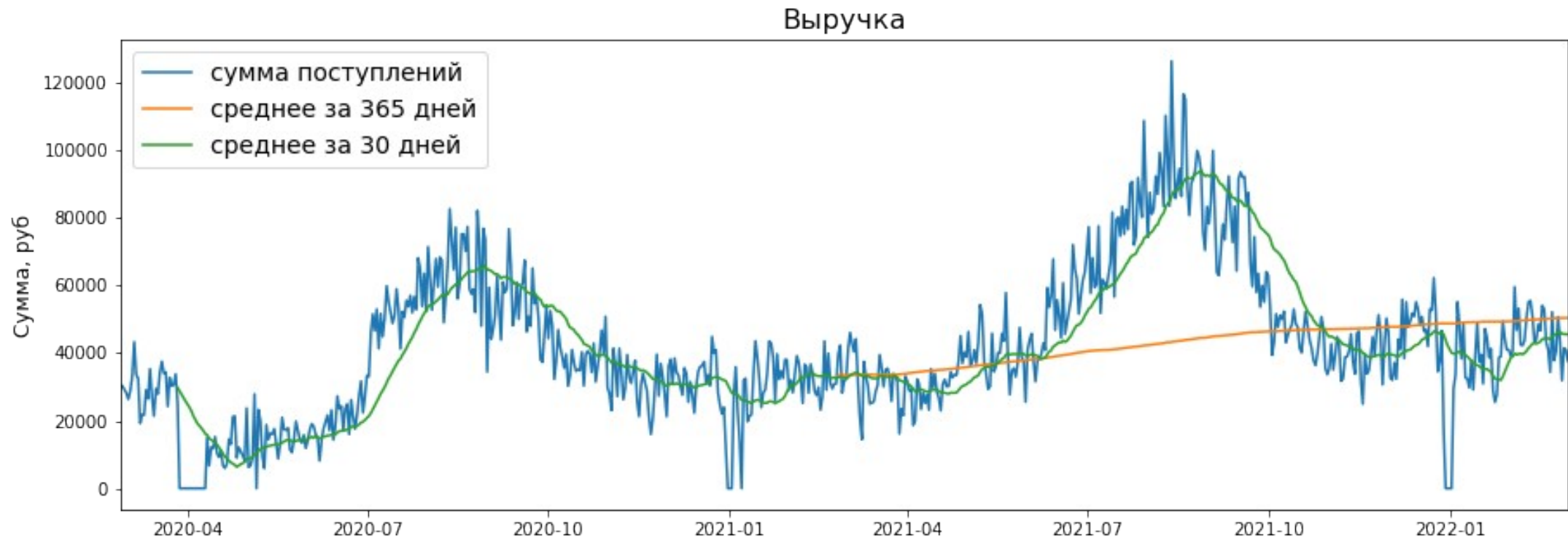


# Статистическая информация

Для анализа выбрана сумма всех поступлений.

Средняя выручка составила 41897 рублей, медианное значение - 38825 рублей, максимальная выручка в день - 126438 рублей.

Наблюдаются еженедельные, ежегодные колебания выручки и небольшой восходящий тренд среднегодовой суммы поступлений.

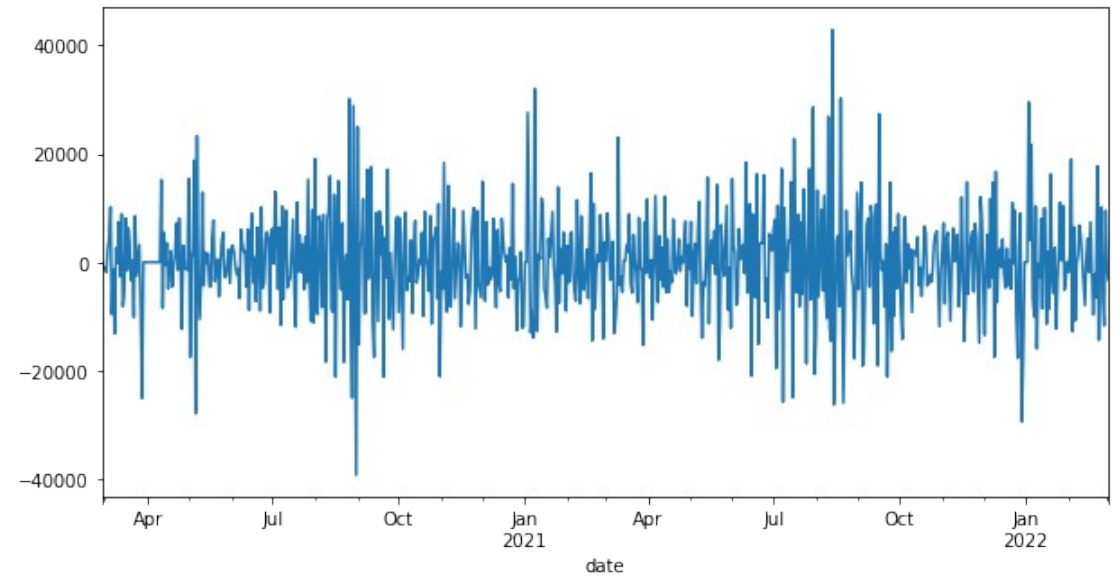


# Оценка параметров временного ряда

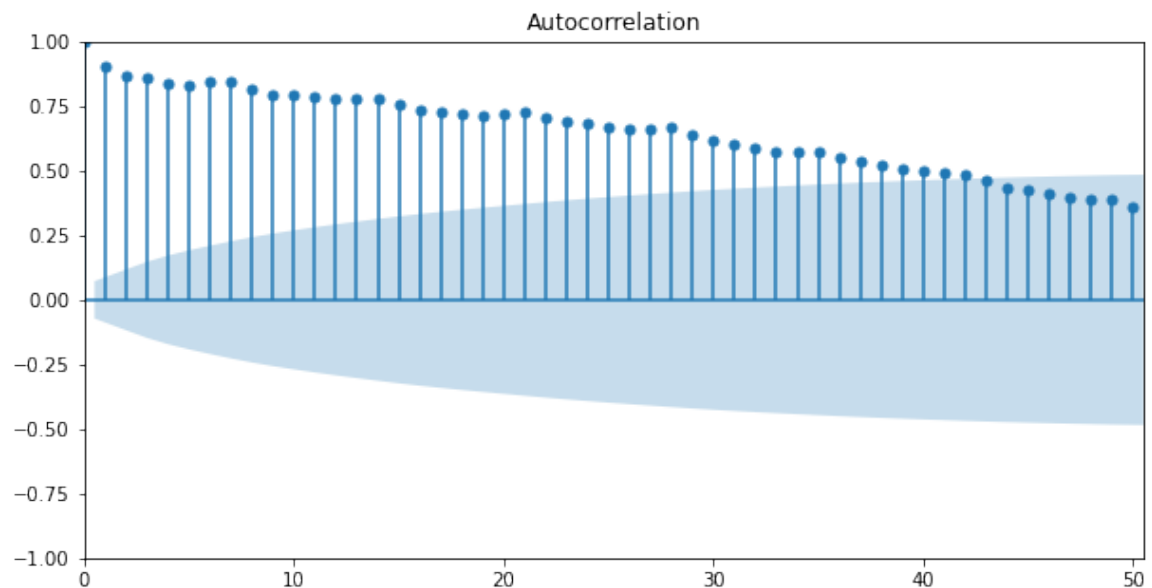
Проведена проверка временного ряда на стационарность с использованием обобщенного теста Дики-Фуллера.

Гипотеза  $H_0$  - временной ряд не стационарен, не отвергается, так как полученное значение  $p\text{-value} = 0.267$ . **Временной ряд нестационарный.**

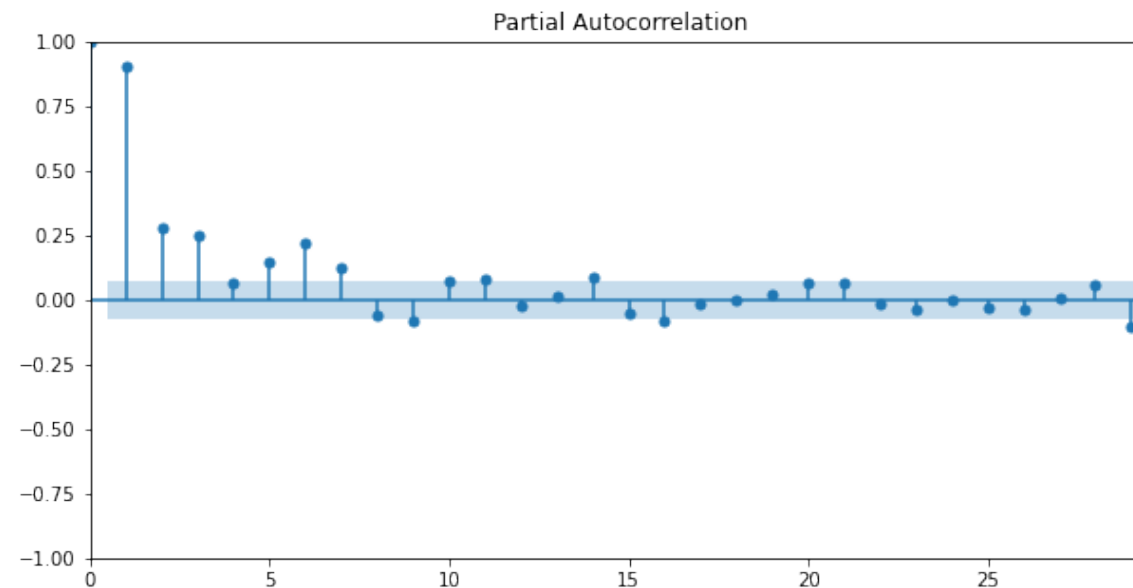
Проведено дифференцирование ряда и проверка полученной первой разности на стационарность. Тест Дики-Фуллера получил  $p\text{-value} = 7.76 \cdot 10^{-12}$ .  
Нестационарность не обнаружена, следовательно наш ряд - **интегрированный ряд первого порядка.**



# Автокорреляционная ACF и частичная PACF функция



По коррелограмме ACF определен коэффициент MA для модели ARIMA,  $q=1$ .



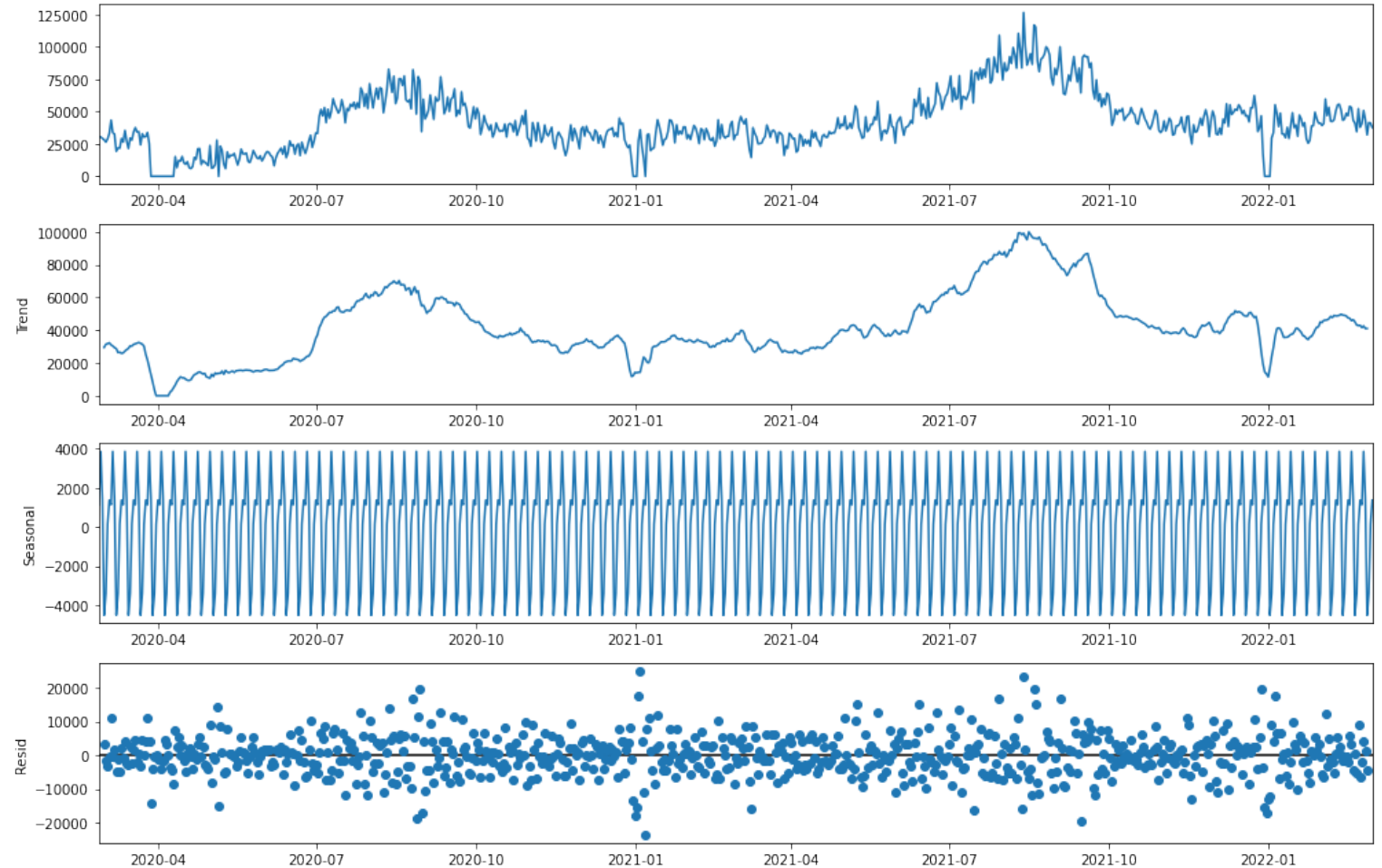
По коррелограмме PACF определен коэффициент AR для модели ARIMA  $p=1$ .

# Декомпозиция временного ряда

Разложение временного ряда на тренд, сезонный и остаточный элементы

На основании графика определена сезонность — 7 дней

Имеется небольшой среднегодовой восходящий тренд и годовая цикличность.



# SARIMA

Авторегрессионное интегрированное скользящее среднее или ARIMA - это метод прогнозирования для одномерных данных временных рядов.

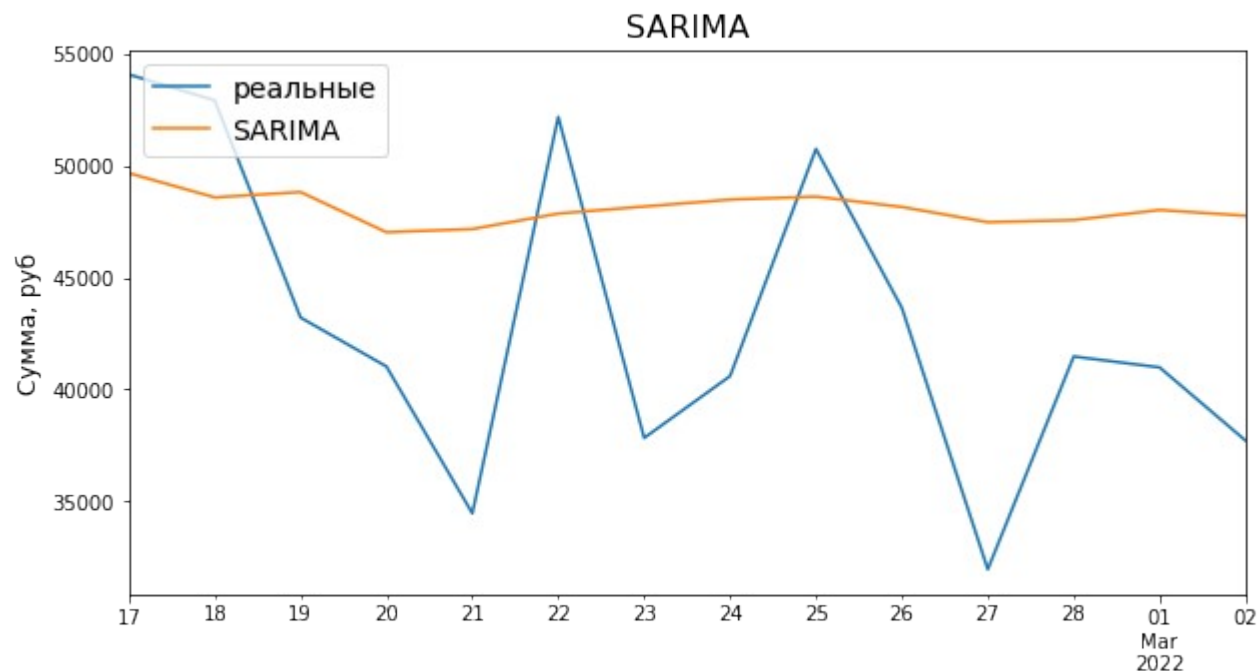
Он поддерживает элементы авторегрессии и скользящего среднего. Интегрированный элемент относится к разнице, позволяющей методу поддерживать данные временных рядов с помощью тренда.

Сезонное авторегрессионное интегрированное скользящее среднее, SARIMA или Seasonal ARIMA, является расширением ARIMA, которое явно поддерживает одномерные данные временных рядов с сезонным компонентом. Использован автоматический подбор параметров модели SARIMA из пакета pdarima.

## SARIMAX Results

Dep. Variable:	y	No. Observations:	721
Model:	SARIMAX(1, 1, 2)x(0, 0, 2, 7)	Log Likelihood	-7514.678
Date:	Sat, 12 Mar 2022	AIC	15041.356
Time:	11:24:27	BIC	15068.832
Sample:	02-27-2020 - 02-16-2022	HQIC	15051.964
Covariance Type:	opg		

MAPE: 0.1841254929  
MSE: 65007278.45  
RMSE: 8062.709126



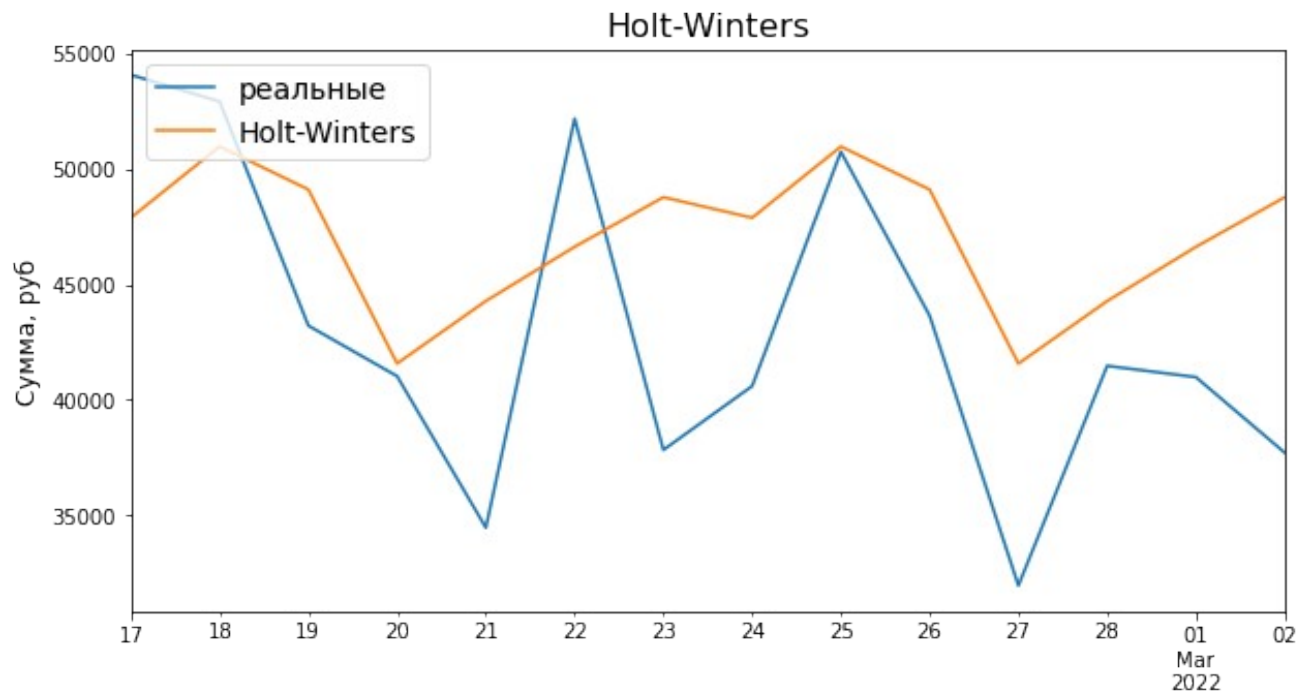
# Тройное экспоненциальное сглаживание

Экспоненциальное сглаживание Холта-Винтерса, названное в честь двух его авторов – Чарльза Холта и Питера Винтерса, является одним из старейших методов анализа временных рядов, который учитывает тенденции и сезонность при прогнозировании.

Этот метод имеет три основных аспекта: среднее значение с учетом тренда и сезонности.

Эти три аспекта представляют собой три типа экспоненциального сглаживания, поэтому метод также известен как тройное экспоненциальное сглаживание.

MAPE: 0.1494773923  
MSE: 47354806.43  
RMSE: 6881.482865

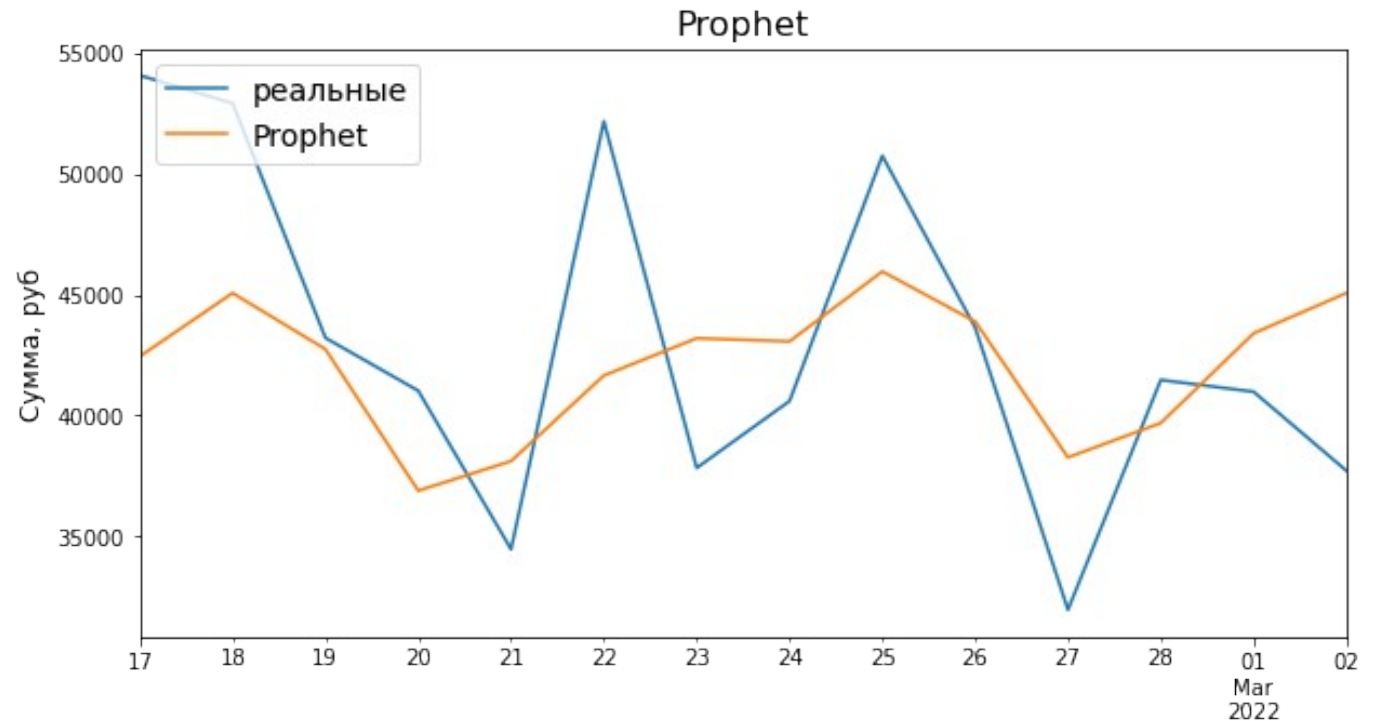




# Prophet

Prophet — библиотека с открытым исходным кодом от компании Facebook. По словам разработчиков (команды Core Data Science team) она хорошо работает с временными рядами, которые имеют ярко выраженные сезонные эффекты, а также имеют несколько таких периодов. Prophet устойчив к отсутствию данных и достаточно хорошо справляется с выбросами. Модель построена с учетом недельной сезонности и годовой цикличности.

MAPE: 0.112958284  
MSE: 35723728.14  
RMSE: 5976.933005



# Выбор модели прогнозирования

Библиотека Prophet показала лучшие результаты прогнозирования на тестовом массиве данных. Этот метод будет использован для построения прогноза выручки на 14 дней.

MAPE:

SARIMA: 0.1841254929

Holt-Winters: 0.1494773923

Prophet: 0.112958284

MSE:

SARIMA: 65007278.45

Holt-Winters: 47354806.43

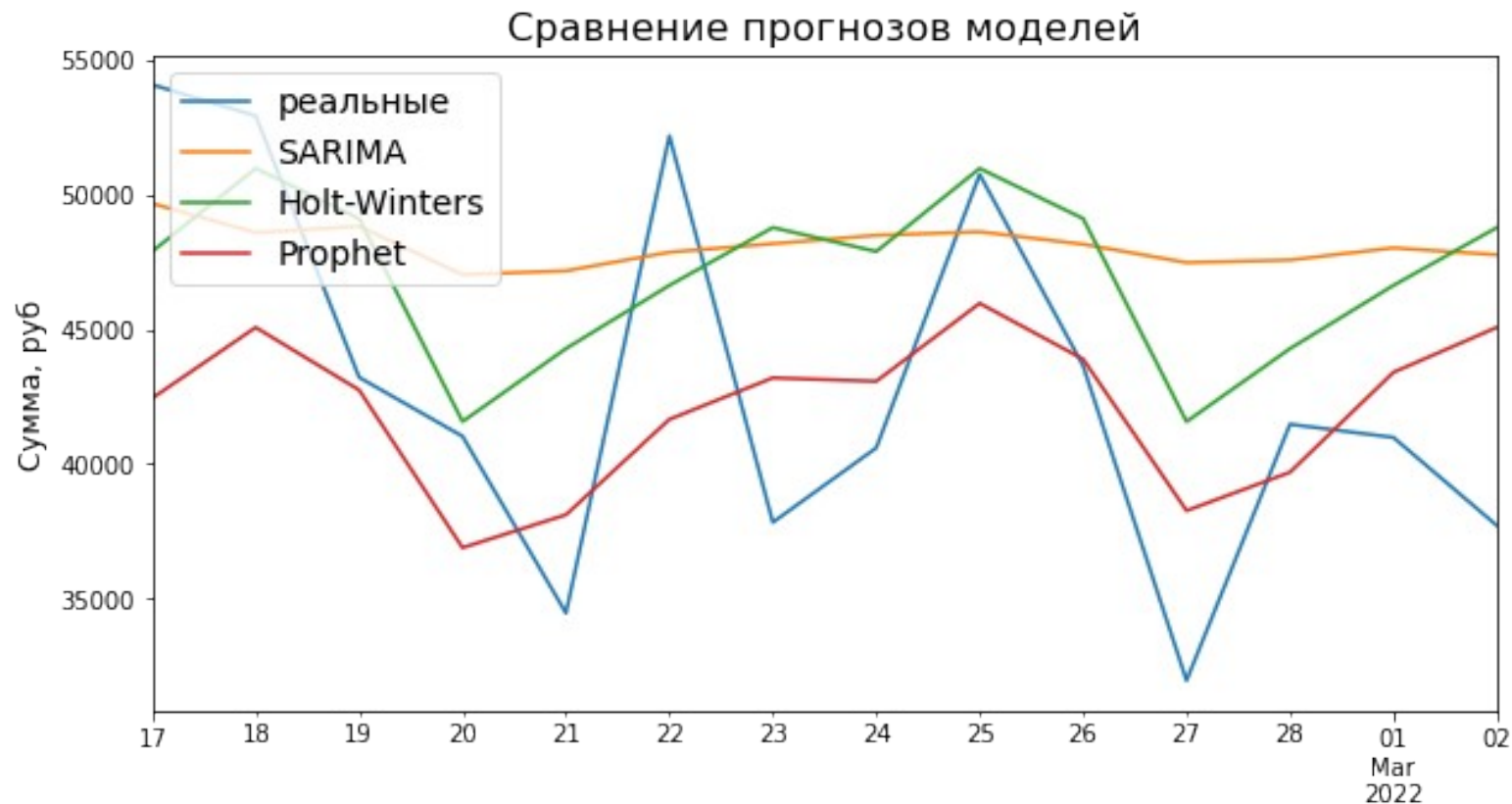
Prophet: 35723728.14

RMSE:

SARIMA: 8062.709126

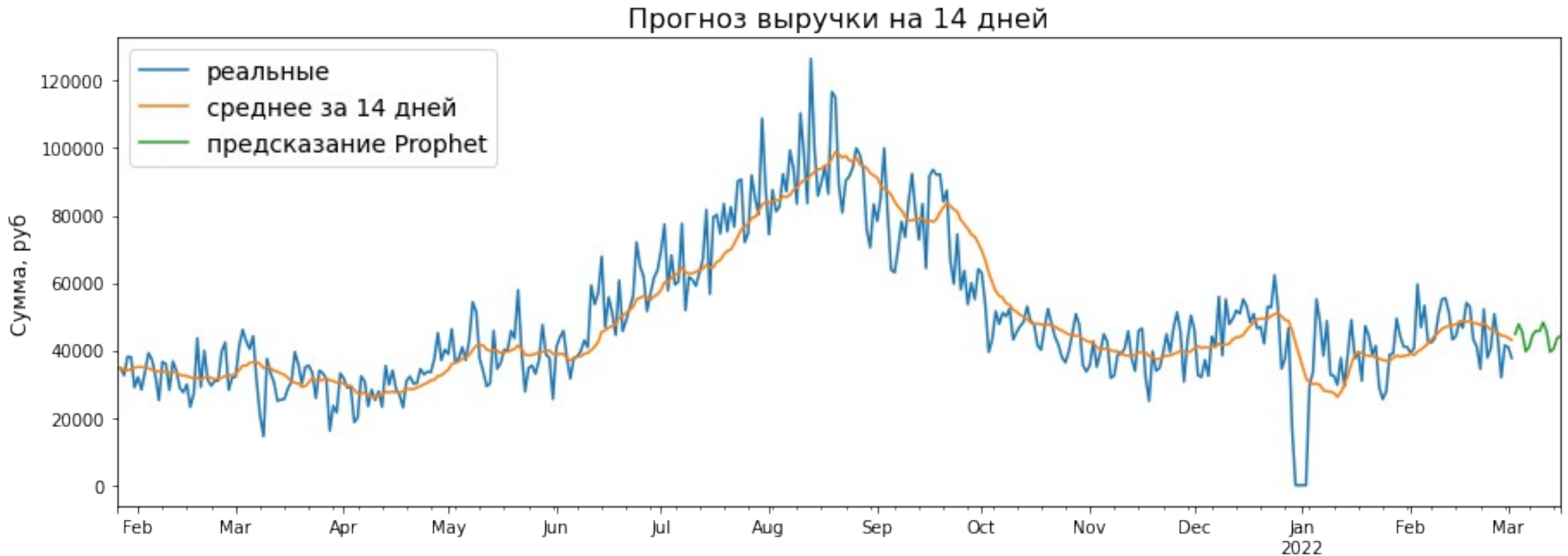
Holt-Winters: 6881.482865

Prophet: 5976.933005



# Прогноз выручки на 14 дней

Произведено прогнозирование с помощью библиотеки Prophet. Предсказание находится на уровне скользящего среднего и учитывает сезонность.



# Выводы

Проведен анализ финансовых поступлений придорожного кафе.

Построен прогноз на предстоящие 14 дней.

Прогнозируемая выручка находится в районе среднегодовых показателей, то есть около 40000 рублей.

Полученный прогноз соответствует ежегодному сезонному колебанию выручки.

На размер выручки влияет множество других факторов, которые нет возможности учесть при анализе.

Это такие как вводимые ограничения при повышении эпидемиологической опасности, открытие придорожных сетевых магазинов Магнит, Пятерочка, Светофор, конкуренция с другими предприятиями общепита, дорожные работы на трассе, ограничивающие транспортный поток или подъезд к кафе, аварийные или плановые отключения электроэнергии.

Также на выручку влияет тип самих посетителей, постоянные или случайные, и сложившееся у них впечатление о качестве блюд и обслуживании персонала.

Все эти неучитываемые факторы, влияющие и на конкурентов, могут оказать на выручку как отрицательное так и положительное действие.

Поэтому построить долгосрочный прогноз исследованными методами не представляется возможным, а к полученным прогнозным данным надо относиться с определенной степенью доверия.

Возможно построение нейронной сети, учитывающей различные факторы от экономических до социальных, позволит давать более точные предсказания.



**Спасибо за внимание.**

[https://github.com/krutofal/homework/blob/main/Forecast cafe.ipynb](https://github.com/krutofal/homework/blob/main/Forecast%20cafe.ipynb)