**Індивідуальне завдання**

Виконав: Колдаєв Дмитро, група КА-03мп

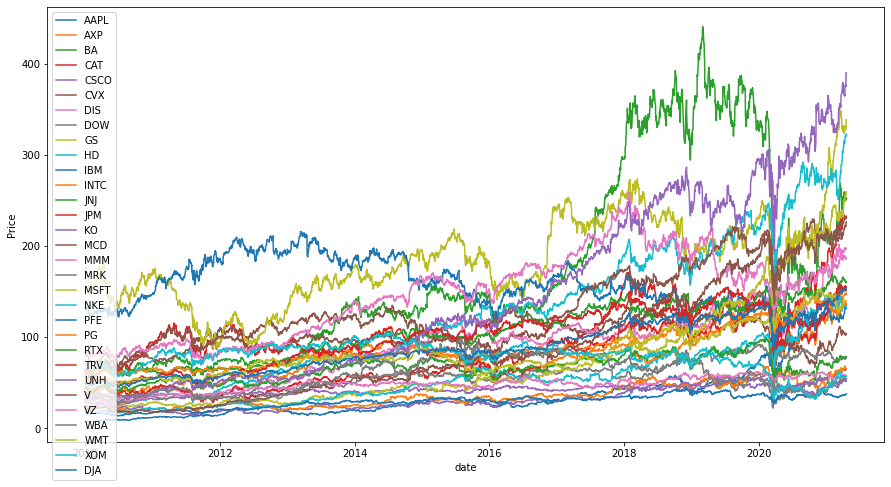
Прогнозування фінансових процесів сучасними методами

Мета: *навчитися застосовувати сучасні методи прогнозування для*

*аналізу та моделювання реальних фінансових процесів, що характеризуються високою волатильністю і потребують побудови моделей з урахуванням їх гетероскедастичності та сезонності.*

**Опис вхідних даних**

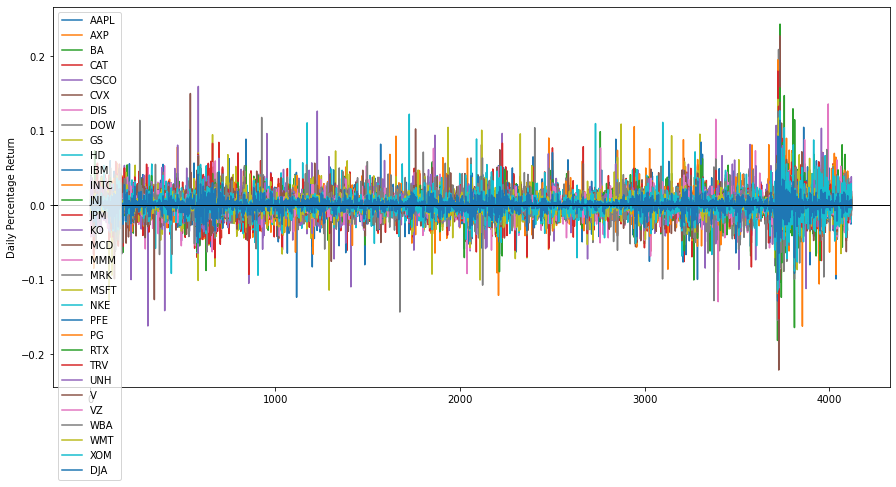
В якості вхідних даних були використані складові індексу Доу-Джонса взяті з платформи yahoo.finance



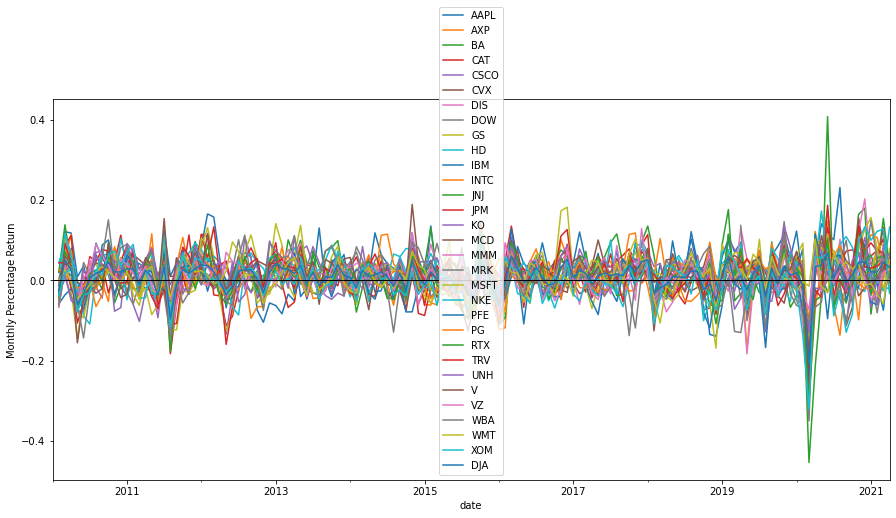
**Експлораторний аналіз**

Для побудови моделей було проведено лінійну інтерполяцію на вихідному наборі даних з метою позбавлення від пропущених значень в цільовій змінній.

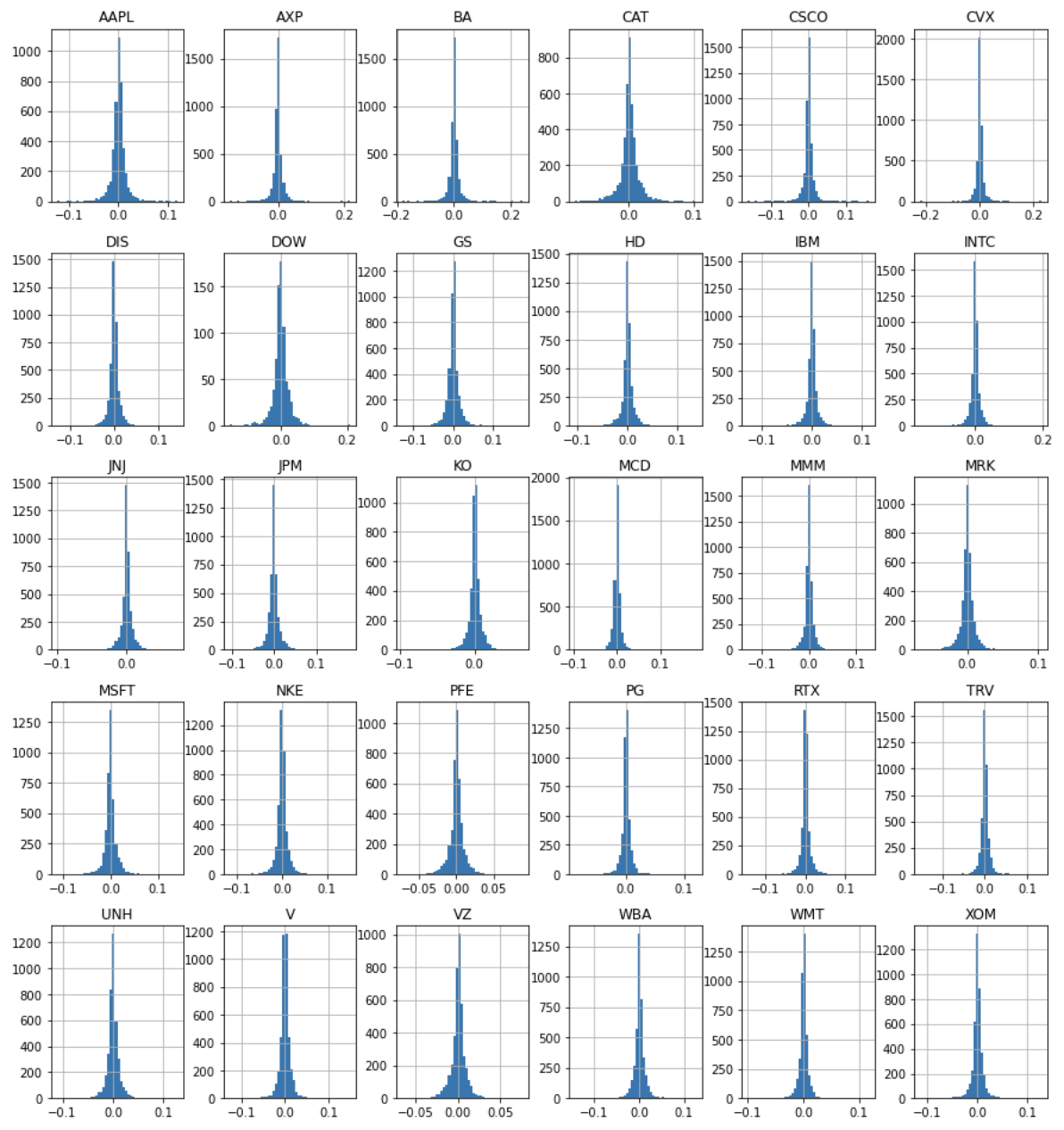
Часові ряди не є стаціонарними, тому розглянемо відносний приріст ціни акцій за день



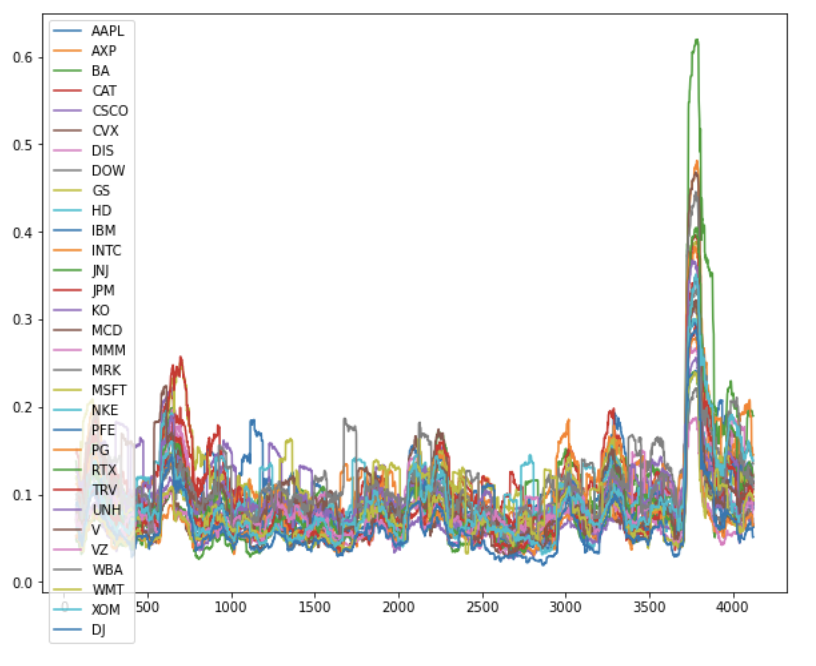
Та приріст за місяць



Розподіли значень приросту відобразимо на гістограмах

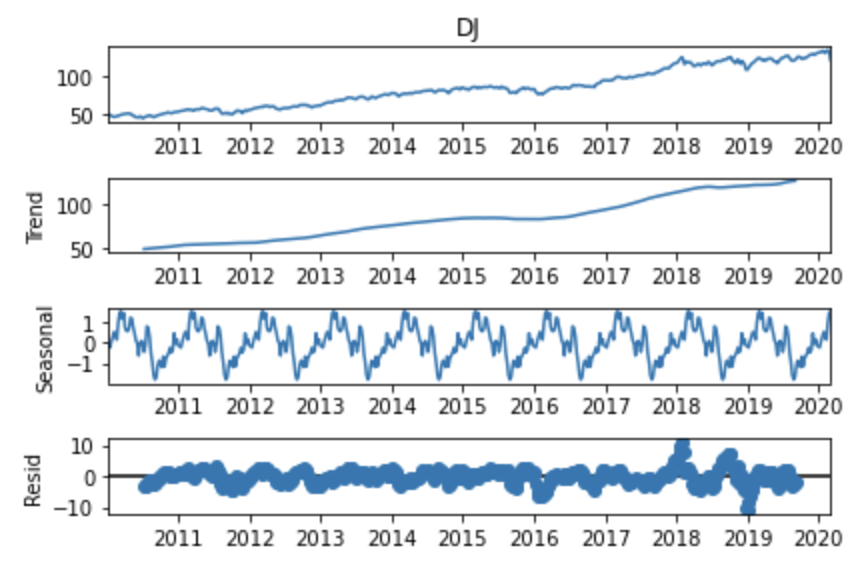
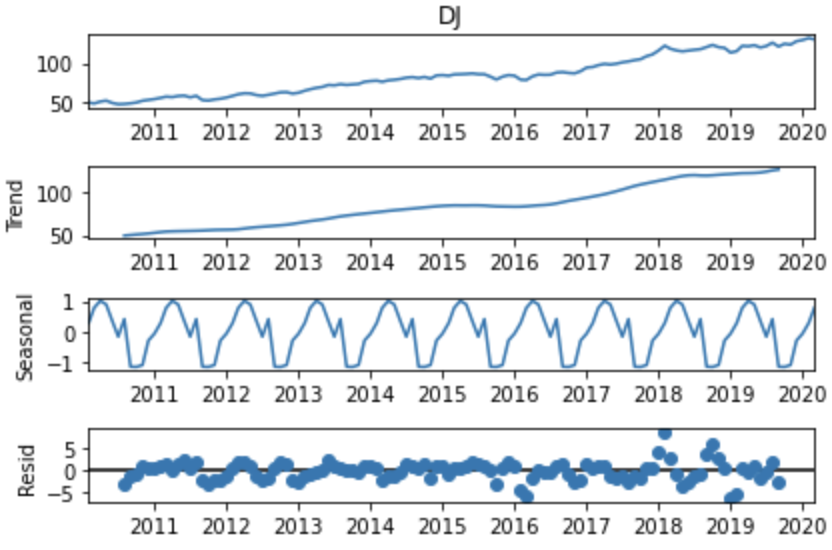


Графіки волатильності часових рядів очікувано аномальні в перший період пандемії COVID-19



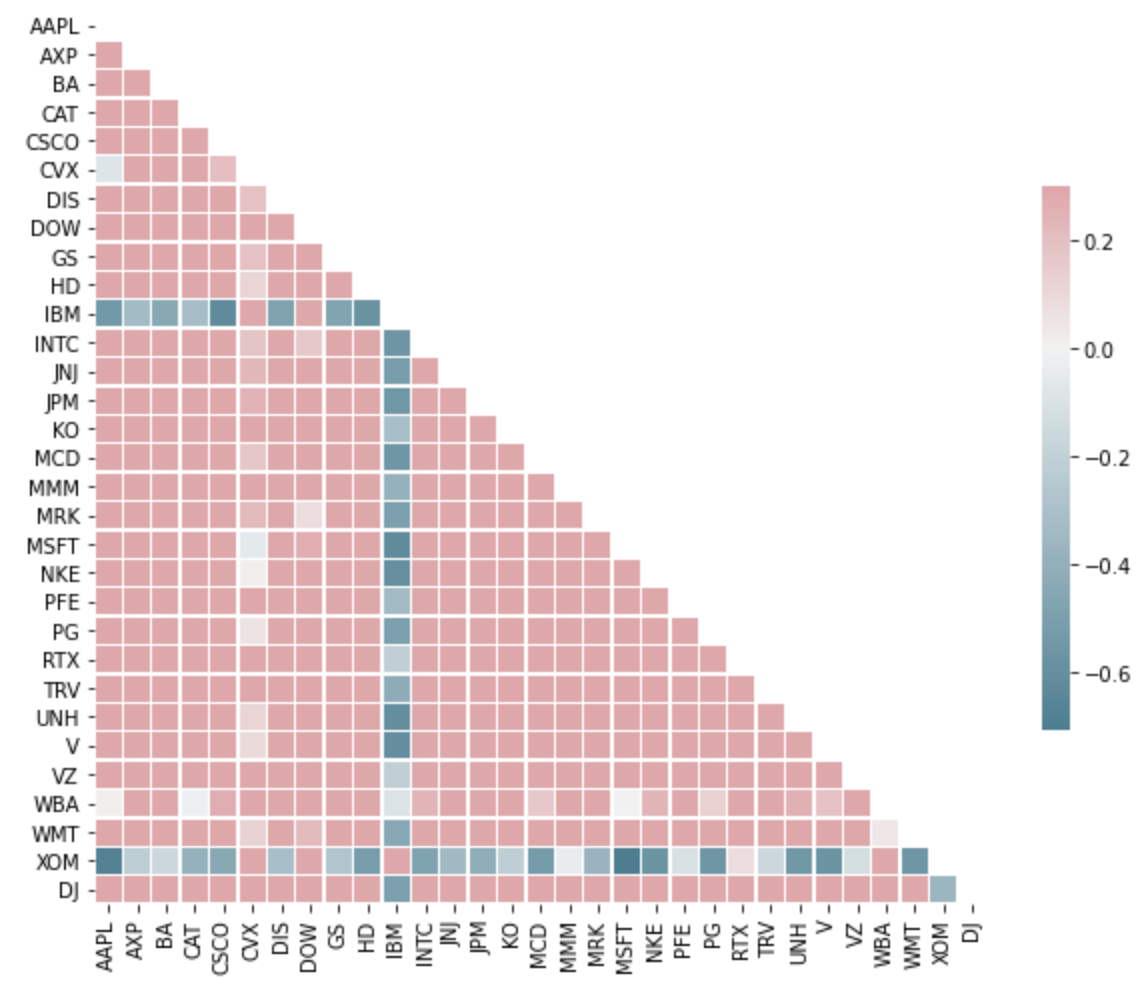
Розглянемо часовий ряд індексу Доу-Джонса. Для цього усереднимо значення по всім акціям що в нього входять

Проведемо аналіз сезонності індексу Доу-Джонса



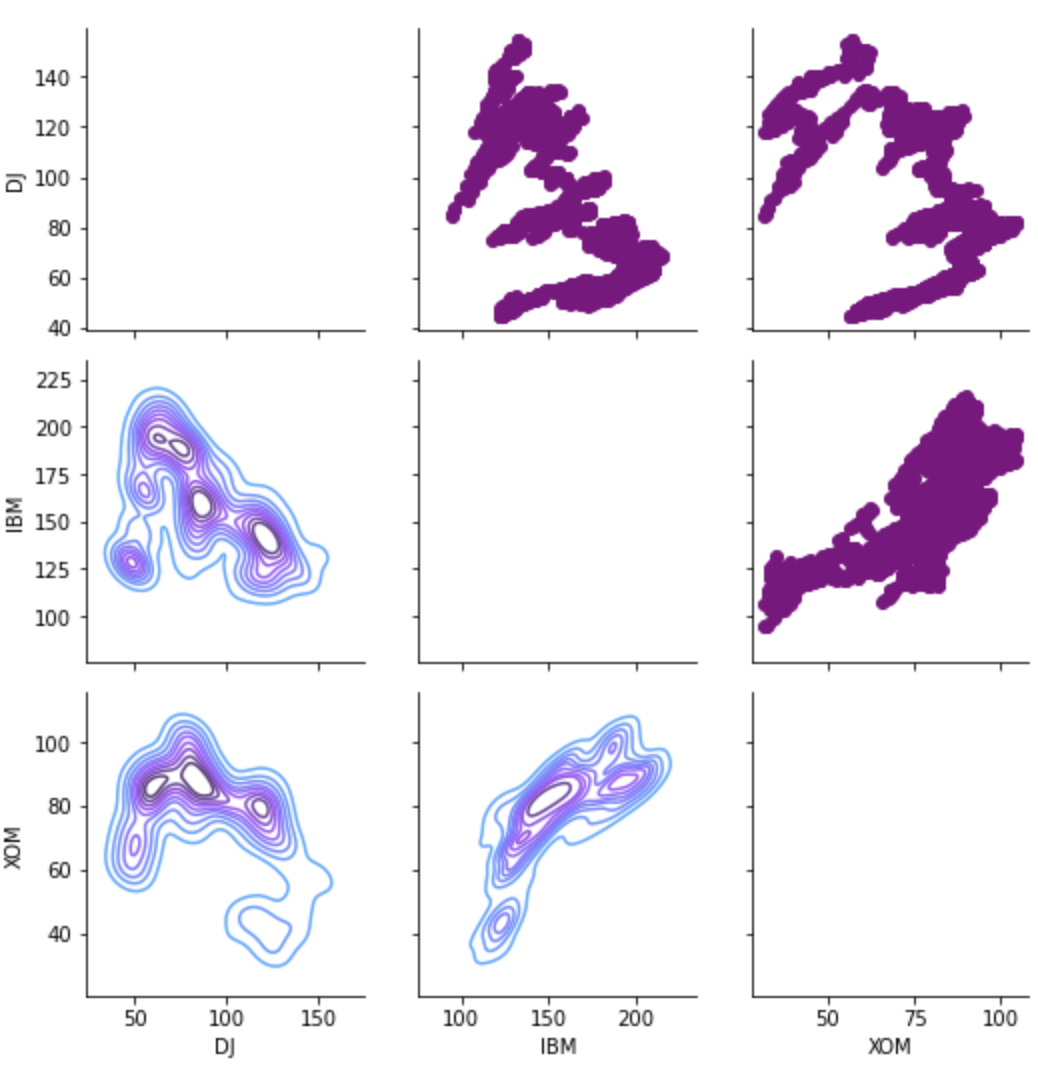
Можемо стверджувати про відсутність сезонності через хаотичний характер залишків

Оцінимо кореляцію між часовими рядами



Між більшістю рядів існує сильна позитивна кореляція, отже доцільно розглядати їх не окремо а в сукупності

Також побудуємо графіки попарну залежність між даними в рядах XOM, IBM, та DJ - де кореляція найменша

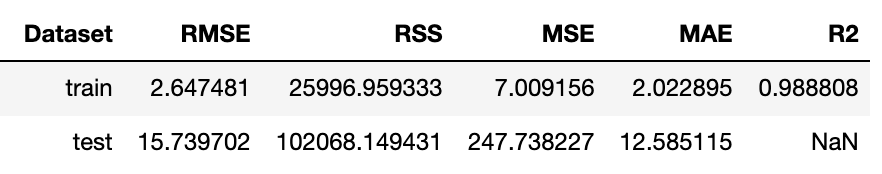


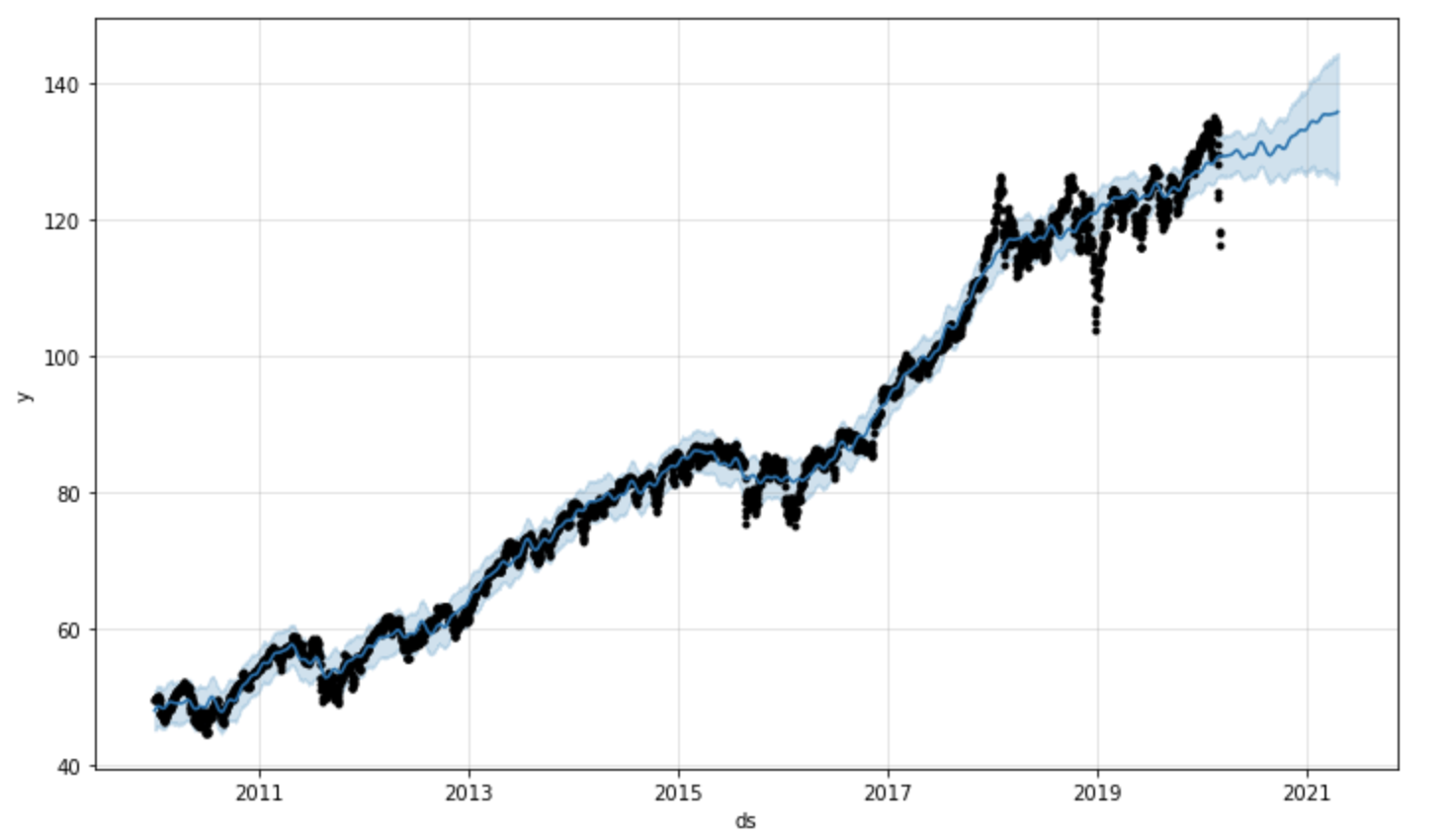
**Прогнозування**

Для прогнозування було вибрано 3 моделі - FB Prophet, LSTM, LightGBM

1. FB Prophet

FB Prophet показав непогані результати на тренувальних даних, але досить погані на тестових





1. LightGBM

Для пошуку оптимальних параметрів бустингу застосуємо grid search

Маємо наступні оптимальні параметри

{'subsample': 1,

'scale\_pos\_weight': 1301,

'num\_leaves': 37,

'num\_iterations': 1000,

'min\_data\_in\_leaf': 1,

'min\_child\_weight': 4.51,

'max\_depth': 9,

'learning\_rate': 0.1,

'lambda\_l2': 0.001,

'lambda\_l1': 0.05,

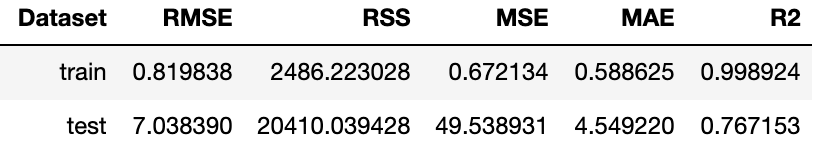
'feature\_fraction': 0.6000000000000001,

'extra\_trees': True,

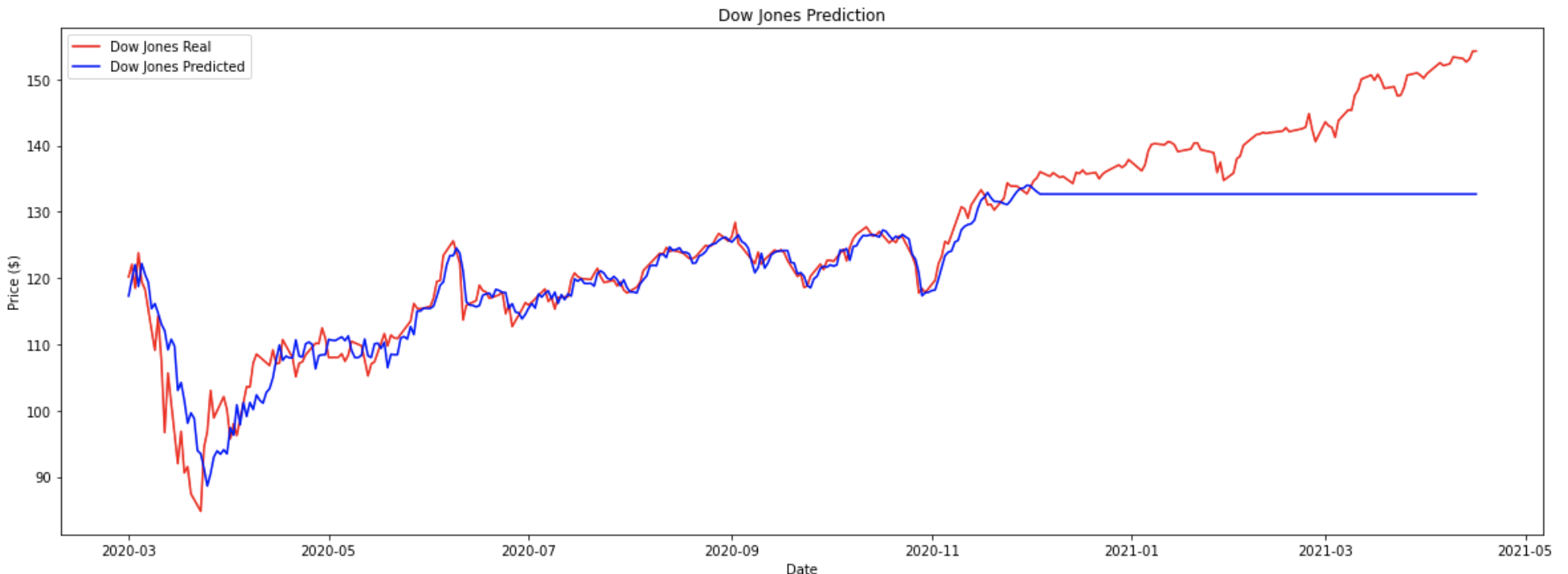
'colsample\_bytree': 0.4,

'bagging\_freq': 1}

З оптимальними параметрами модель дала непогані результати



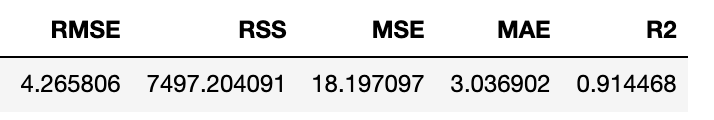
Недоліком є майже статичний прогноз останніх місяців тестової вибірки

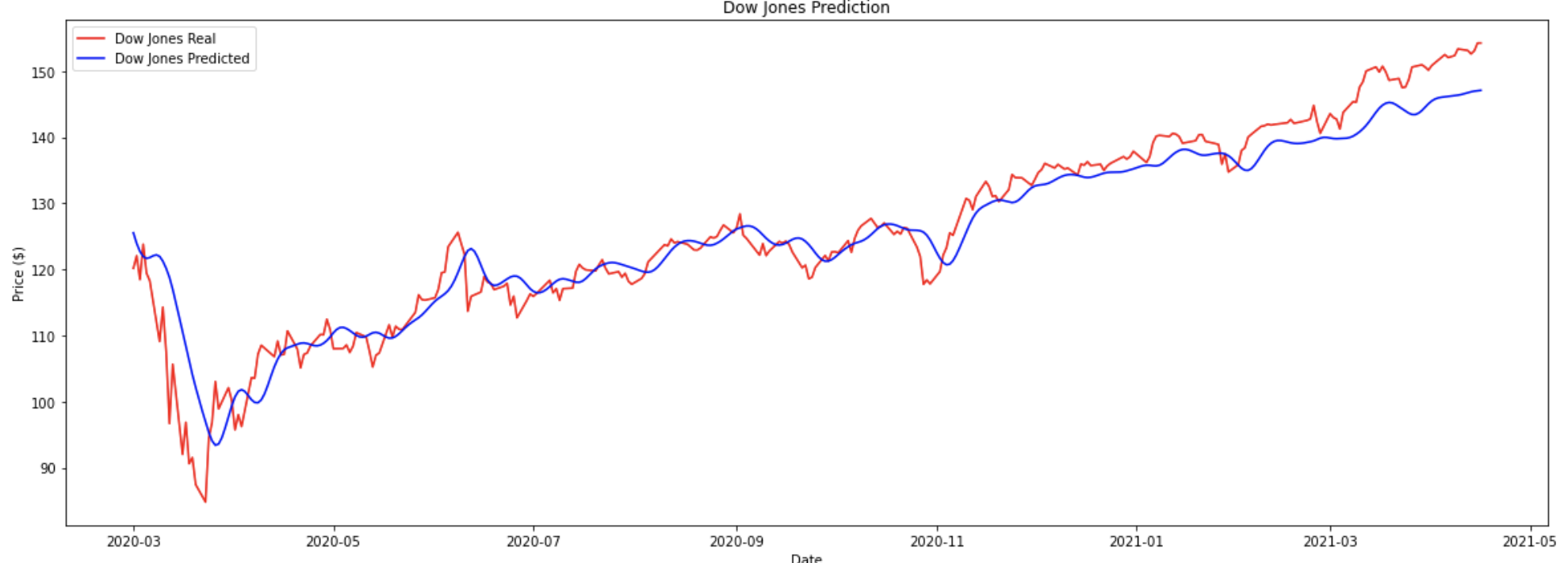


1. LSTM

Модель LSTM будуємо з 4 однакових шарів по 30 нейронів і параметром дропауту 0.2 - ці параметри були обрані експертно

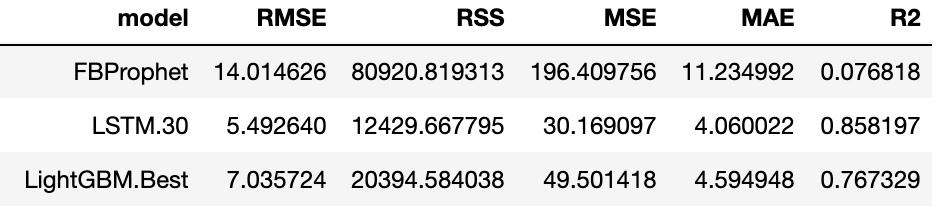
Дана модель на тестовій вибірці показала наступні метрики



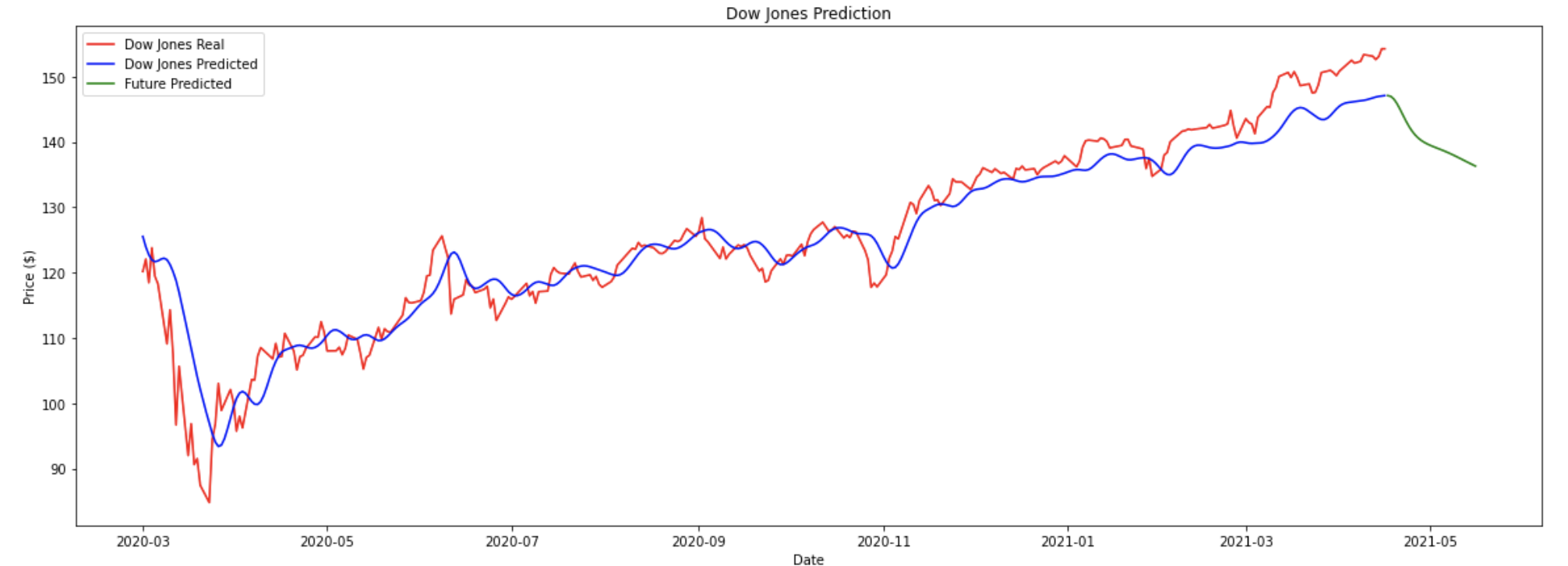


1. Вибір кращої моделі

Отже кращою моделлю за порівнянням на тестовій вибірці виявилаь модель LSTM



Розглянемо також динамічний прогноз цієї моделі в майбутнє



**Висновки**

Для побудови кращої моделі для прогнозування індексу Доу-Джонса було проведено комплексний експлораторний аналіз та очистка даних. Далі було застосовано 3 різних підходи до прогнозування ціни - за допомогою рішення від Фесбук, бустингу на нейронної мережі. Найкращі результати показала нейронна мережа LSTM