**Міністерство освіти і науки України**

**Національний Технічний Університет України**

**«Київський політехнічний інститут ім. Ігоря Сікорського»**

**Теплоенергетичний факультет**

**Кафедра автоматизації проектування енергетичних процесів і систем**

**Лабораторна робота №3**

з дисципліни

«Прикладні методи аналізу даних»

Виконав:

аспірант групи ТВ-71ф

Ляшенко М.М.

**Київ - 2018**

### Обраний набір даних

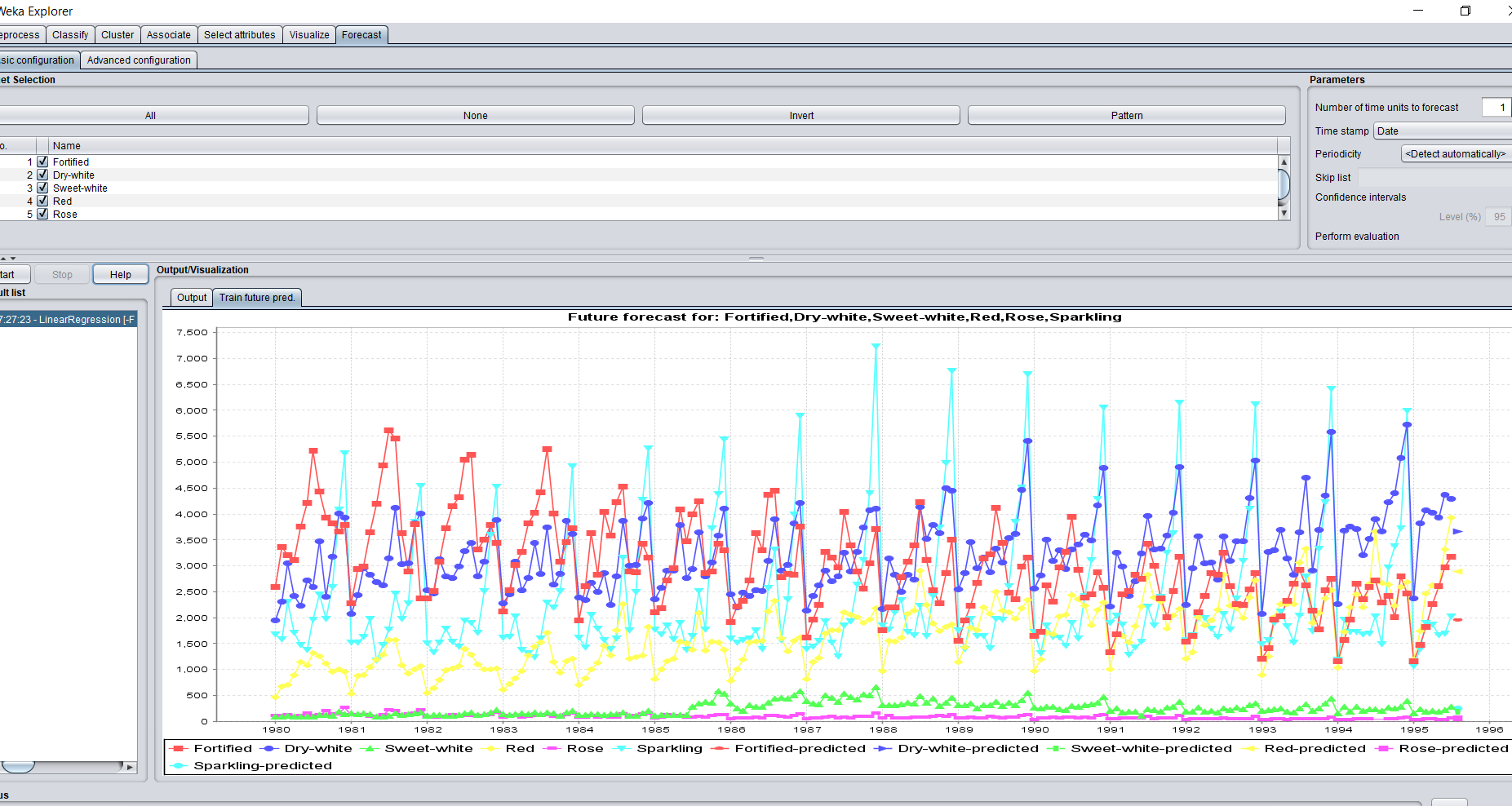
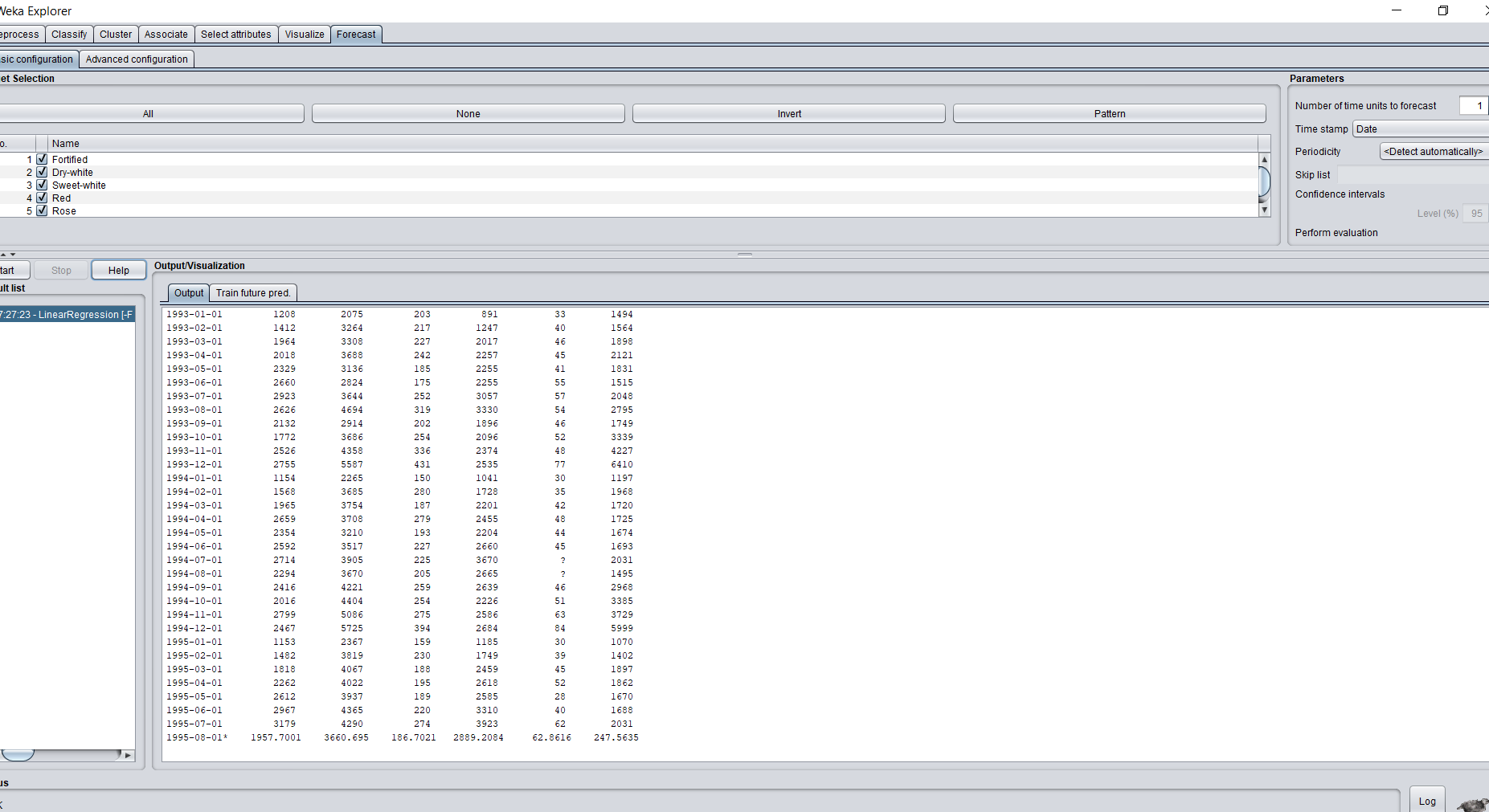
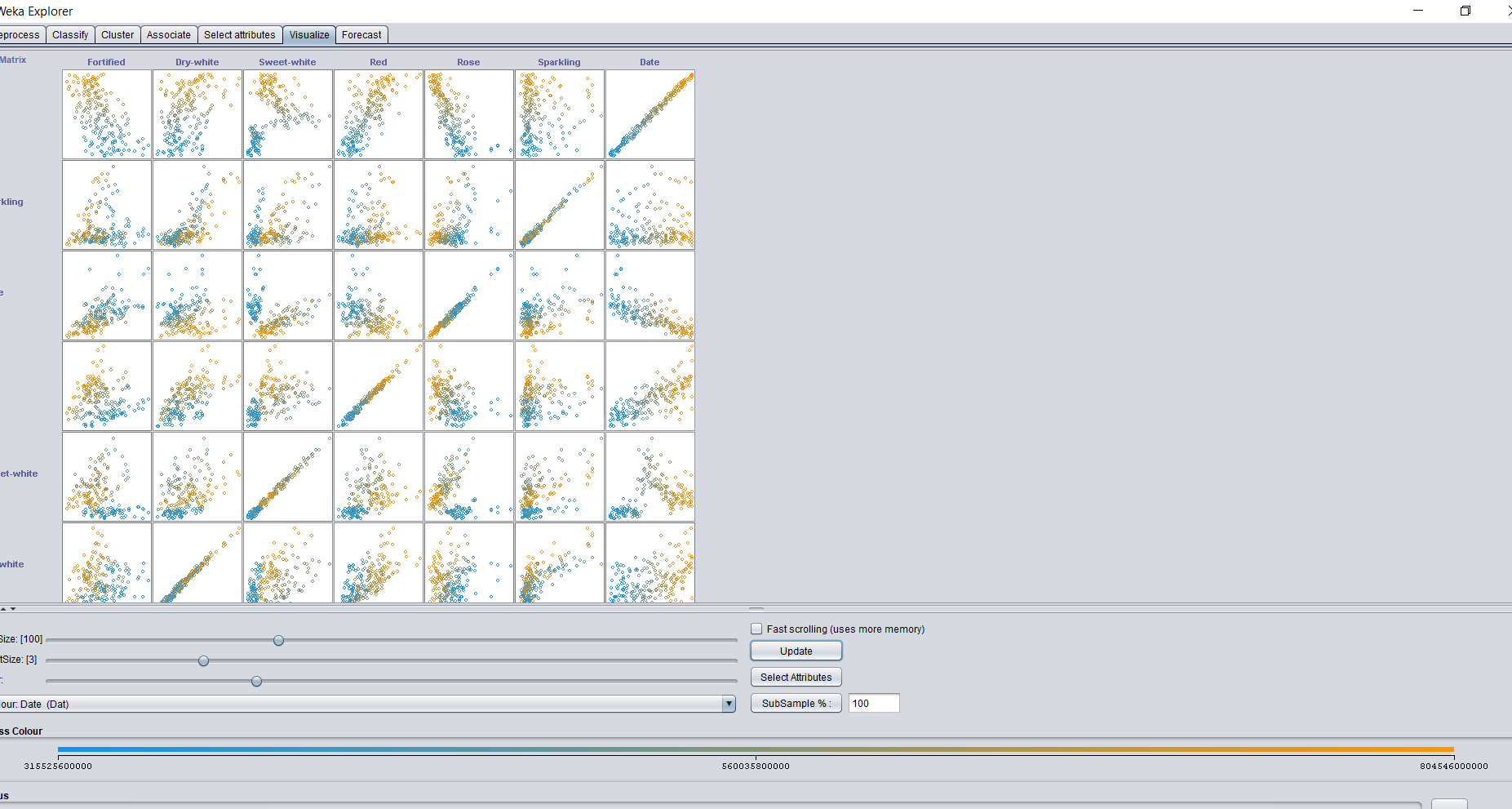
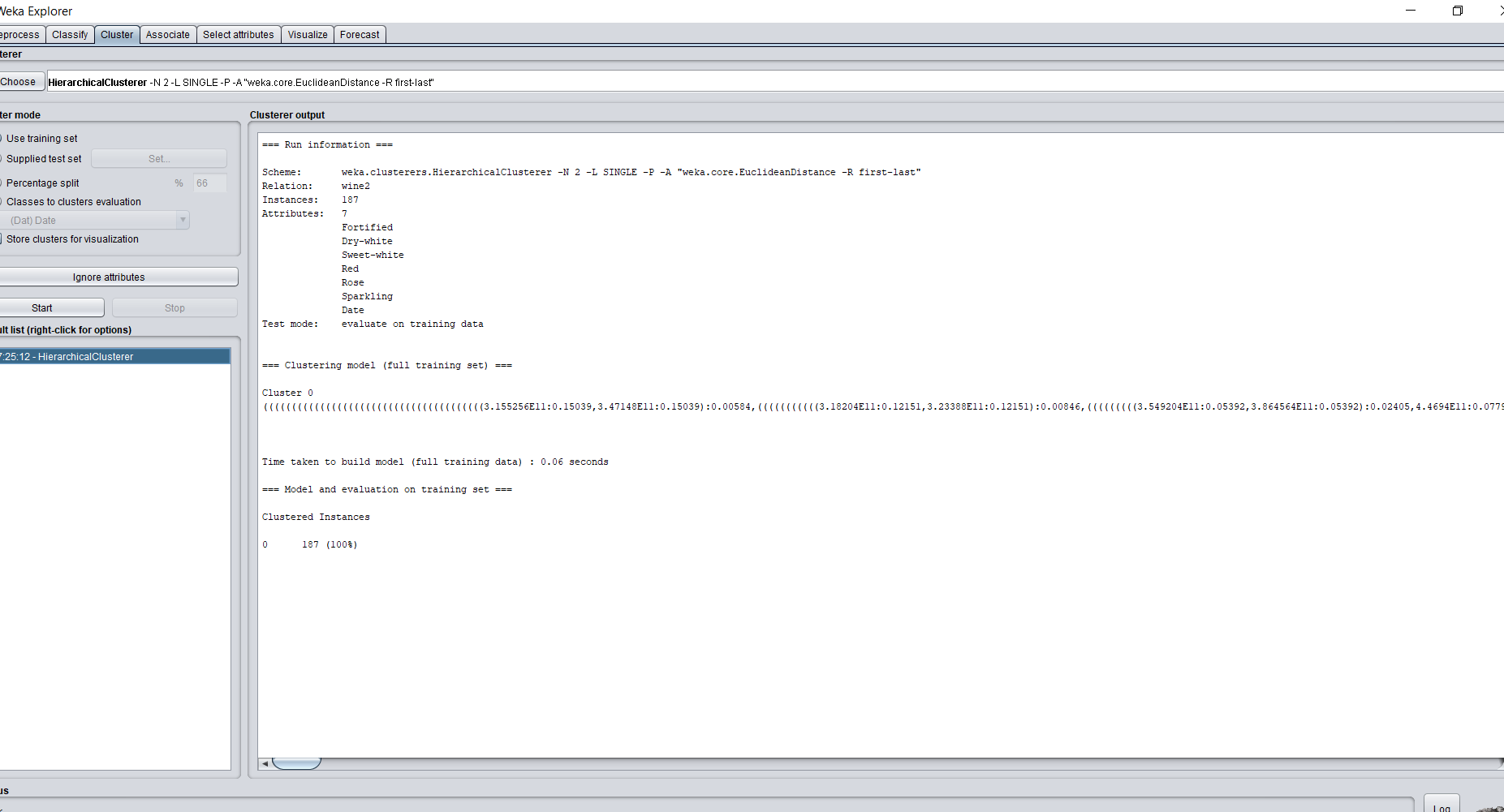
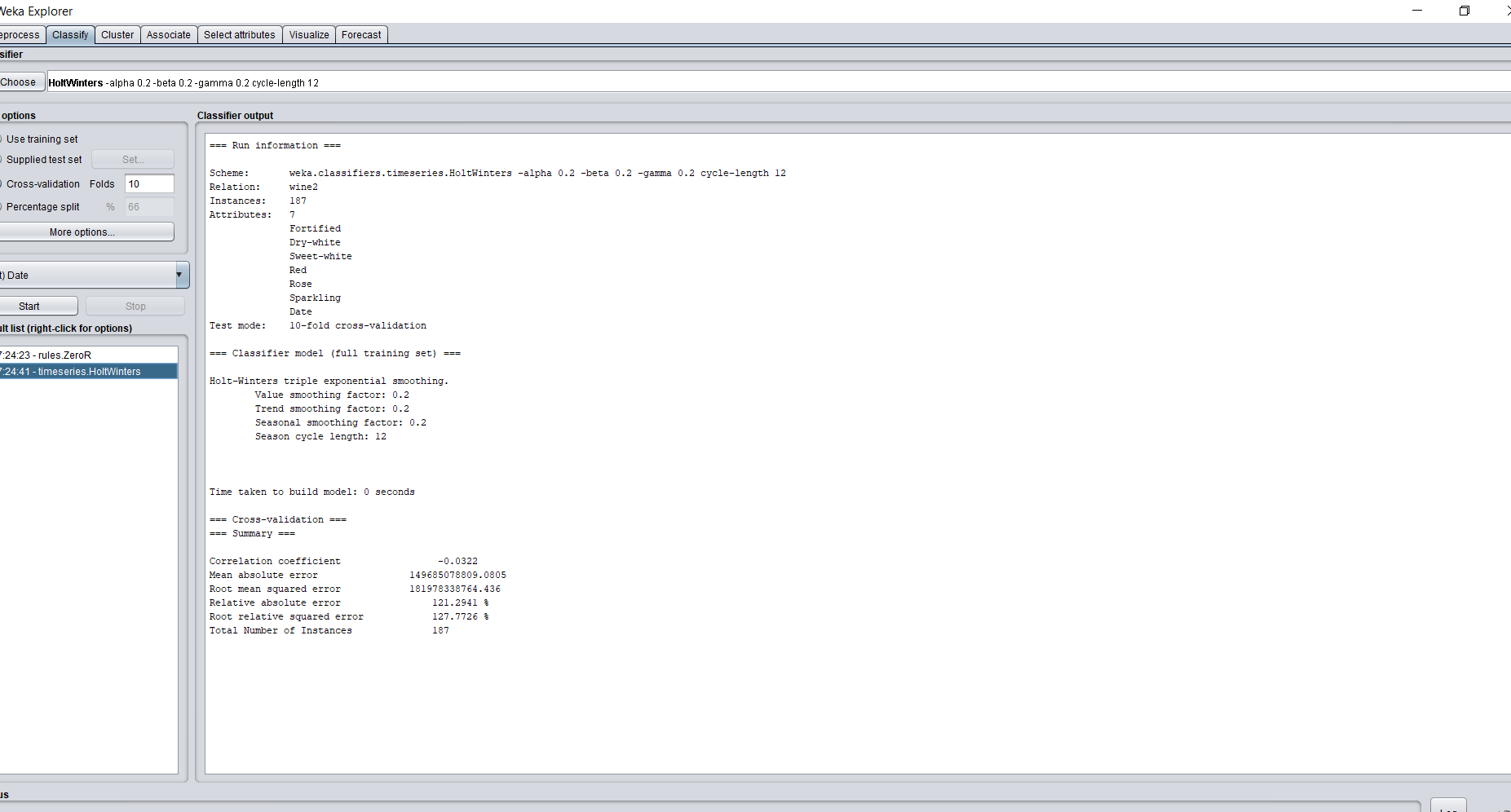
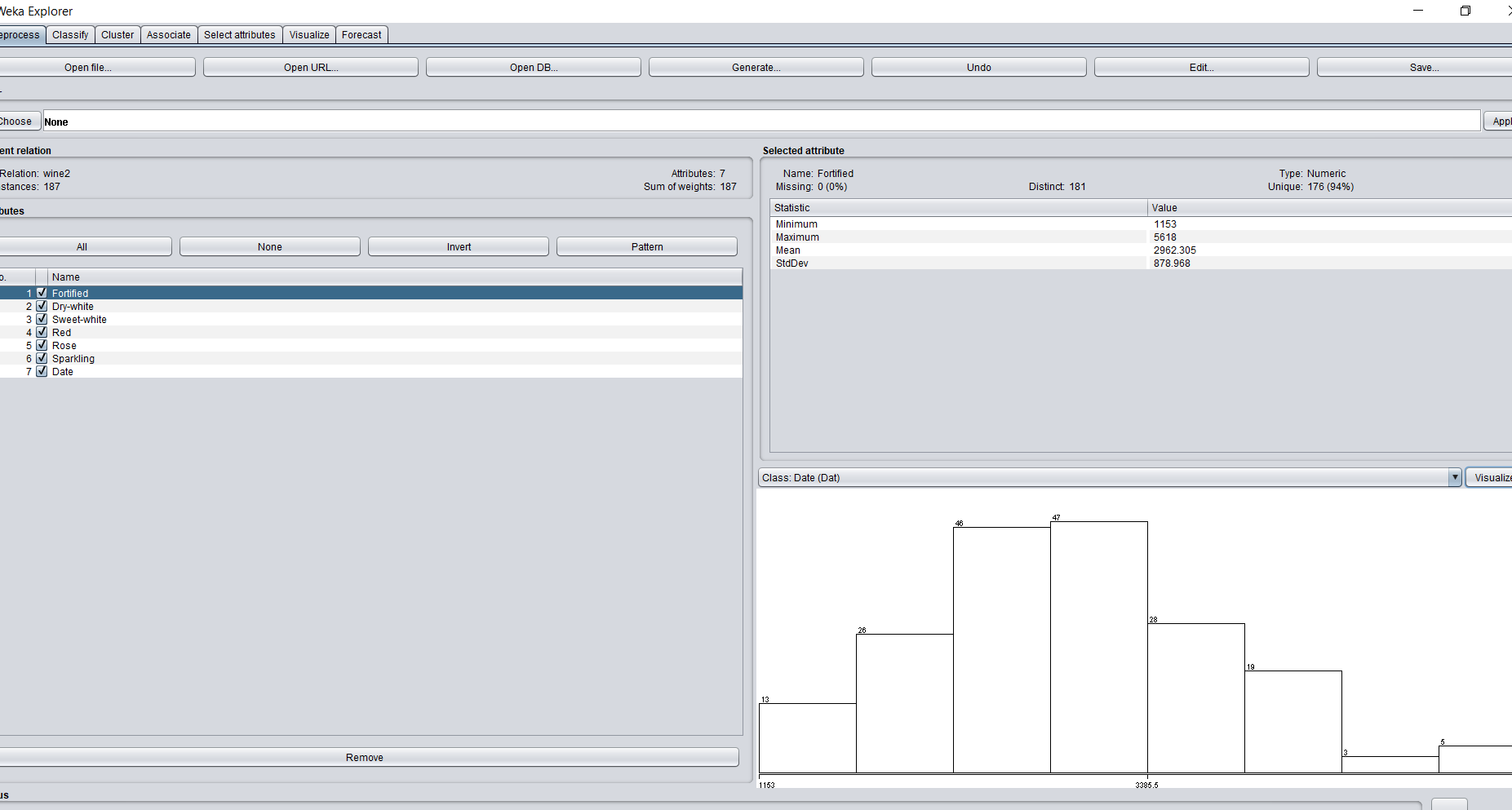
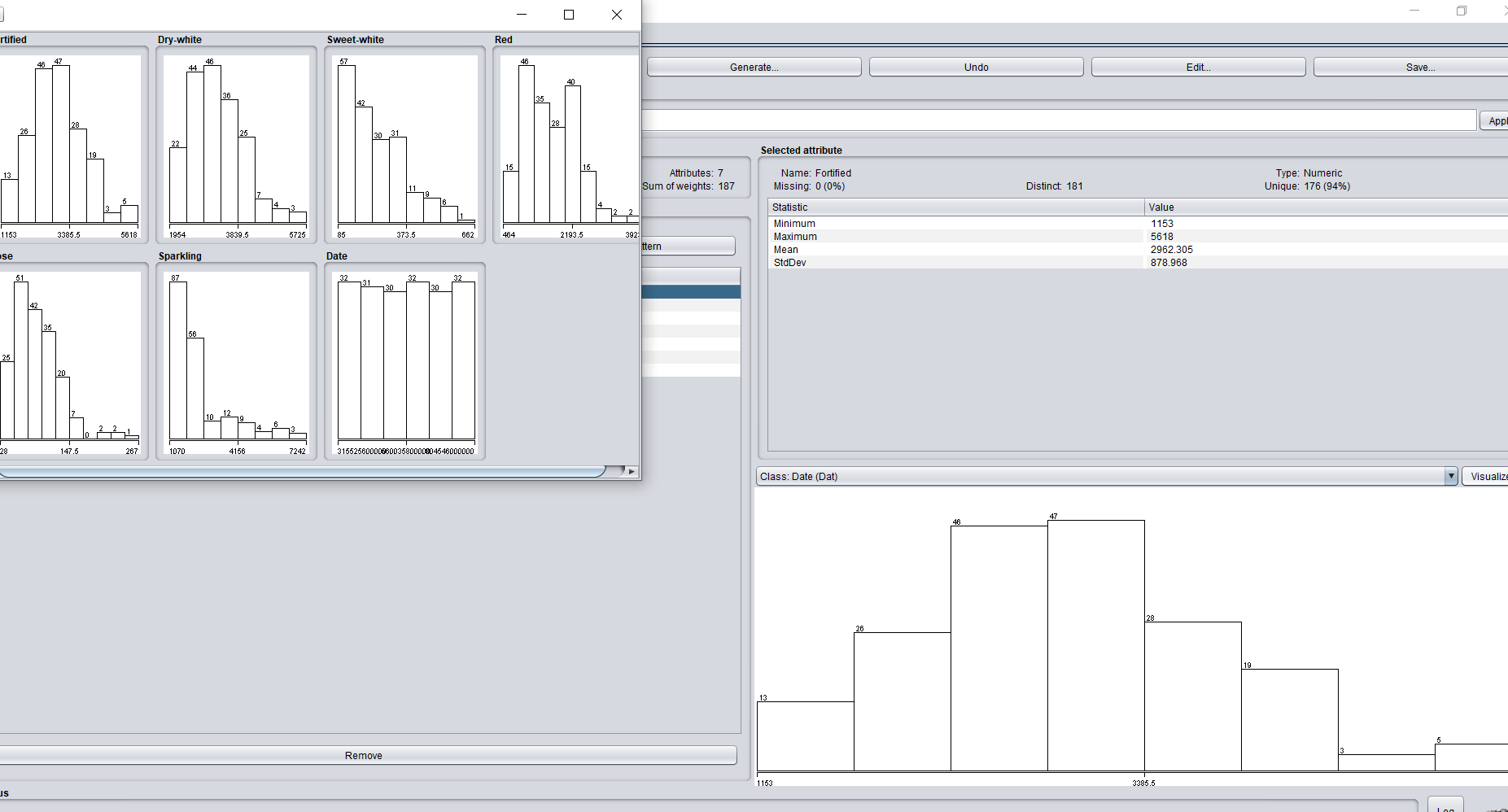
Набір даних було взято звідси:

<http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka>

Аналіз часових рядів - це процес використання статистичних методів для моделювання та пояснення залежних від часу рядів даних. Прогнозування часових рядів - це процес використання моделі для створення прогнозів (прогнозів) для майбутніх подій на основі відомих минулих подій. Дані часових рядів мають природне тимчасове впорядкування - це відрізняється від типових програм інтелектуального аналізу даних / машинного навчання, де кожна точка даних є незалежним прикладом концепції, і впорядкування точок даних у наборі даних не має значення. Прикладами додатків часових рядів є: планування потужності, поповнення запасів, прогнозування продажів і майбутній рівень персоналу.

Weka тепер має спеціальне середовище аналізу часових рядів, що дозволяє розробляти, оцінювати і візуалізувати моделі прогнозування. Це середовище має вигляд вкладки плагінів у графічному інтерфейсі користувача "Weka" і може бути встановлено за допомогою менеджера пакетів. Структура часових рядів Weka приймає підхід до машинного навчання / інтелектуального аналізу даних для моделювання часових рядів шляхом перетворення даних у форму, яку можуть обробляти стандартні алгоритми навчання. Це виконується шляхом видалення тимчасового впорядкування окремих прикладів введення, кодуючи залежність від часу через додаткові поля введення. Ці поля іноді називають "відсталими" змінними. Різні інші поля також обчислюються автоматично, щоб дозволити алгоритмам моделювати тенденції і сезонність. Після перетворення даних, будь-який з алгоритмів регресії Weka може бути застосований для вивчення моделі. Очевидним вибором є застосування множинної лінійної регресії, але може бути застосований будь-який метод, який може передбачати безперервну ціль - у тому числі потужні нелінійні методи, такі як векторні машини підтримки для регресійних і модельних дерев (дерева рішень з функціями лінійної регресії на листках) . Такий підхід до аналізу та прогнозування часових рядів часто є більш потужним і більш гнучким, ніж класичні статистичні методи, такі як ARMA і ARIMA.

**Результати:**

****

### Висновок

Всі матераіли можна подивитися тут - <https://github.com/maxml/MachineLearning>

Для приклдау були взяти продажі вина в Австралії (тисячі літрів) з 1980 по 1995 рік. Одним з висновків по графікам мое служити те, що кількість проданого ивна не особливо коливалася в ці роки. На початку періоду більше купляли sparkling (яке потім майже повністю перестали купляти) та sweet-white (яке менш повільнішими темпами, але почало продаватися менше), в наступному промідку більше продавалося rose (яке ще довго потім куплялося, але в результаті теж втратило свої позиції), red (це був пік, але пізніше воно також користувалося попитом), dry-white, fortified потім стали основними типами вина. Класифікація ніякого результату не дала, кластеризація також не дала результатів.Серед результатів візуалізації я також не зміг віднайти якогось результату, кожен з них не схоже, що залежний один від одного. Найбільш цікавий результат дала вкладка Forecast – dry-white, red найкраще продавалися в наступному році. На жаль, можливості це перевірити нема.