Міністерство освіти і науки України

НТУУ «Київський політехнічний інститут»

Фізико-технічний інститут

Спеціальні розділи програмування

Лабораторна робота №1

## Наука про дані: підготовчий етап

**Виконав:**

Студент 2 курсу ФТІ

групи ФІ-42

Ластовський О. В.

**Перевірив:**

Колотій А.В.

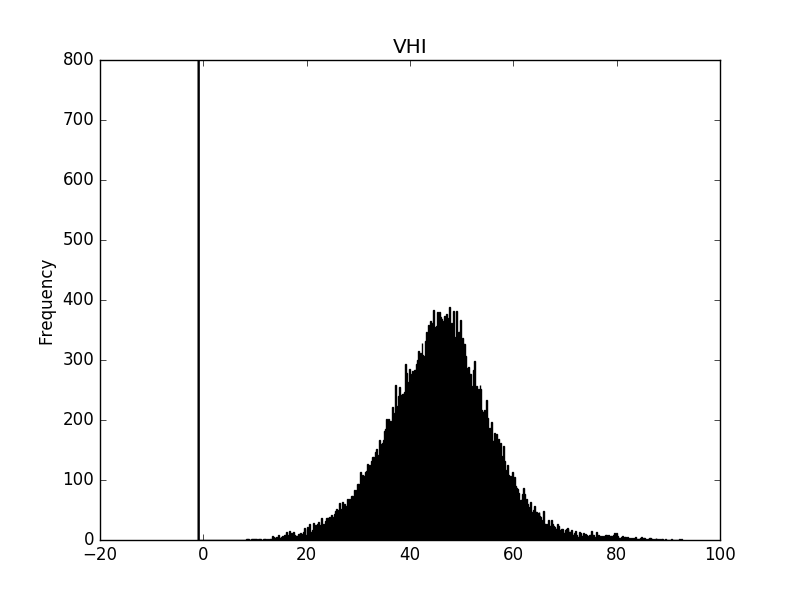
Київ 2016

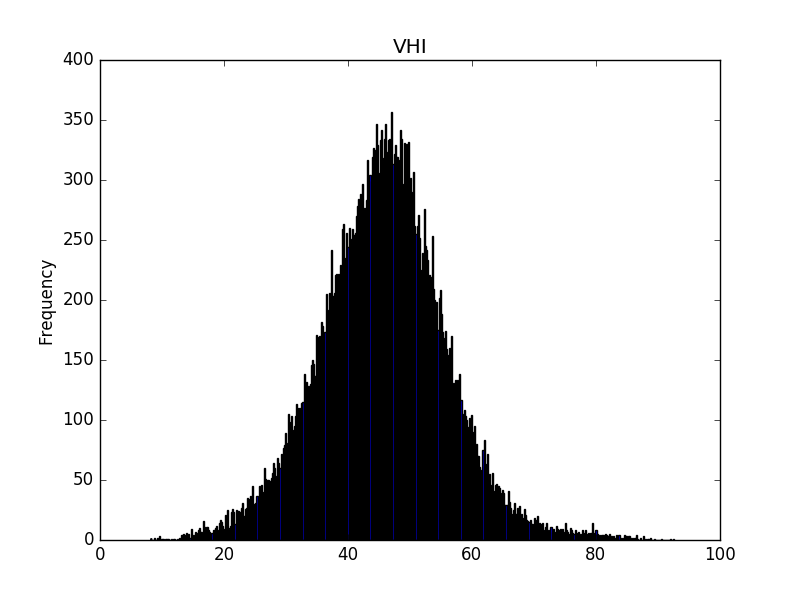
Мета роботи:

Ознайомитися з основними кроками по роботі з даними – workflow від постановки задачі до написання пояснювальної записки, зрозуміти постановку задачі та природу даних, над якими виконується аналітичні операції.

Хід роботи:

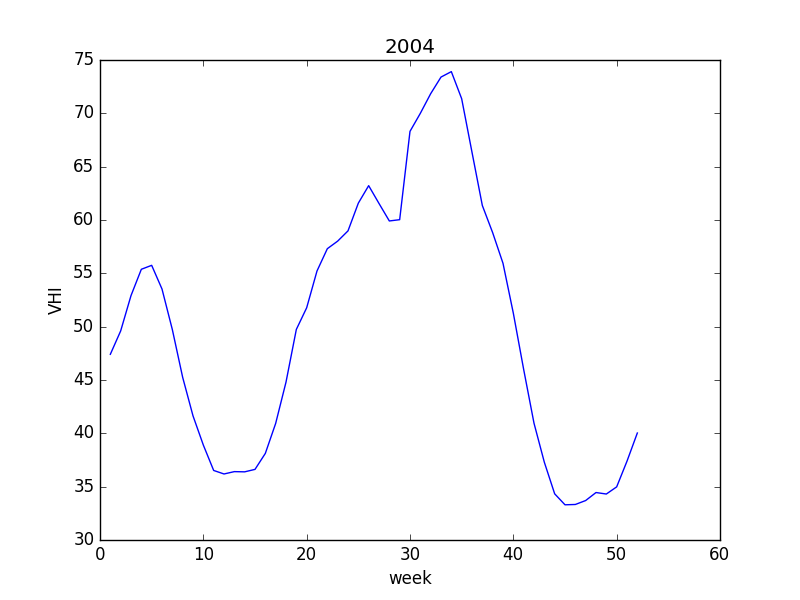
Кількість вхідних даних: 44825 (рядків) x 10 (колонок) = 448250 значень. Масив вхідних даних містить поля, що не мають достовірних значень. Ці поля мають значення -1. Для очистки фрейму від цих даних була реалізована відповідна процедура. Нижче представлена частотна діаграма значень індексу VHI до та після очистки даних:

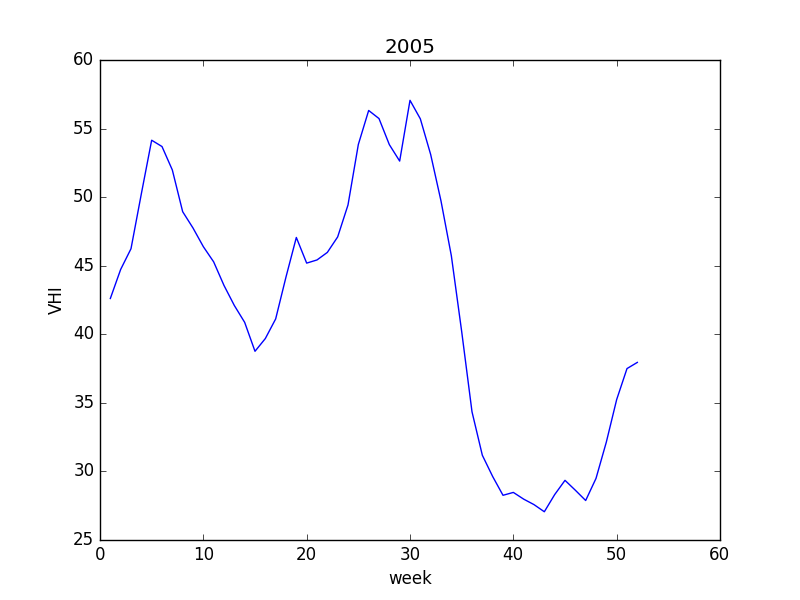




Кількість даних після очистки: 43987 (рядків) x 10 (колонок) = 439870 значень.

Графік залежності індексу VHI від часу впродовж 2004-2005 років для Київської області:





Починаючи з вересня вегетаційний індекс падає до кінця року, із початком зими росте і в кінця січня набуває проміжного максимуму, до початку квітня індекс знову спадає, тоді як в кінці липня – початку серпня досягає свого максимуму.

Висновок:

В ході виконання лабораторної роботи були реалізовані відповідні процедури за заданим алгоритмом. Символи, що позначають відсутність даних були визначені за допомогою аналізу частотних діаграм отриманих даних. Можна виділити такі етапи аналізу даних: отримання даних, обробка, аналіз та інтерпретація результатів обробки.