ชื่อชุดข้อมูล : **Water Quality (Drinking water potability)**

แหล่งที่มาของข้อมูล : [**https://www.kaggle.com/adityakadiwal/water-potability**](https://www.kaggle.com/adityakadiwal/water-potability)

**Why is it interesting?**

น้ำเป็นสิ่งที่จำเป็นต่อการดำรงชีวิตของมนุษย์เป็นอย่างมาก ซึ่งคุณภาพของน้ำเป็น สิ่งจำเป็นที่เราต้องรู้ก่อนที่จะนำเข้าสู่ร่างกาย ไม่ว่าจะเป็นความเป็นกรด-เบส ปริมาณสาร แขวนลอยต่างๆ และ Organic Carbon ที่ถูกสร้างจากแบคทีเรียที่อยู่ในแหล่งน้ำ สิ่งเหล่านี้ล้วนส่งผลต่อคุณภาพน้ำและส่งผลต่อผู้บริโภคเป็นอย่างมาก

**Graphical user interface, text

Description automatically generated with medium confidenceData frame Info :**

**หมายเหตุ :** จะเห็นว่าลำดับของข้อมูลไม่ได้ถูกเรียงตามลำดับเนื่องจากมีการตัดแถวของข้อมูลที่ไม่ทราบค่าออกไป [ na.omit(data) : in R programing ] เพื่อให้ได้การวิเคราะห์ผลที่แม่นยำที่สุดจากข้อมูลทั้งหมด 3276 ตัวอย่างเมื่อตัดข้อมูลที่ไม่ทราบค่าทั้งหมดออกไปจะเหลือข้อมูลทั้งหมด 1980(pH), 2011(Turbidity) ตัวอย่าง และ **เลือกมาวิเคราะห์เฉพาะคอลัมน์ที่สนใจคือ pH, Turbidity และ Potability**

**Fundamental Statistical Value :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Statistical value** | **Non-Drinkable Water** | | **Drinkable Water** | |
| **pH** | **Turbidity (NTU)** | **pH** | **Turbidity (NTU)** |
| **Mean** | 7.0672 | 3.9552 | 7.1138 | 3.9913 |
| **Median** | 6.9920 | 3.9441 | 7.0465 | 4.0073 |
| **Mode** | 8.3168 | 4.6288 | 9.4451 | 3.8752 |
| **1st Quartile** | 5.9829 | 3.4447 | 6.2560 | 3.4406 |
| **3rd Quartile** | 8.1420 | 4.4975 | 7.9552 | 4.5275 |
| **Interquartile** | 2.1591 | 1.0528 | 1.6991 | 1.0868 |
| **Min** | 1.4318 | 1.4500 | 0.2275 | 1.4922 |
| **Max** | 14.0000 | 6.4947 | 11.8981 | 6.4942 |
| **Range** | 12.5682 | 5.0447 | 11.6706 | 5.0020 |
| **STD.** | 1.6591 | 0.7829 | 1.4376 | 0.7764 |
| **Variance** | 2.7526 | 0.6131 | 2.0668 | 0.6028 |

**Chart, histogram

Description automatically generatedChart, histogram

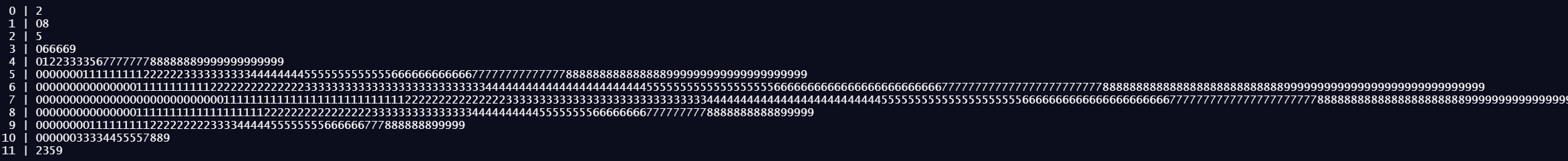
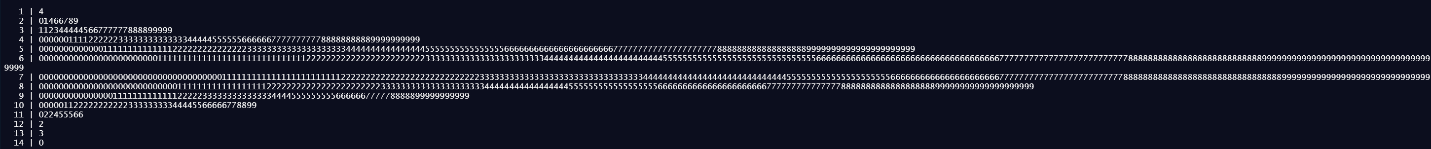
Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generatedHistogram :**

**หมายเหตุ :** NTU ย่อมาจาก Nephelometric turbidity unit โดยเป็นหน่วยวัดความขุ่นของน้ำโดยวัดจากสารแขวนลอยในน้ำในหน่วย mg/l หรือ ppm ซึ่งสารแขวนลอย 1 mg/l เท่ากับ 3 NTU องค์การอนามัยโลกกำหนดไว้ว่าน้ำที่สามารถนำมาบริโภคได้ไม่ควรจะมีค่าของ Turbidity เกิน 5 NTU

**Stem and Leaf :**

* **pH**
  + **Non-Drinkable Water**
  + **Drinkable Water**
* **Turbidity**
  + **A black screen with white text

    Description automatically generated with low confidenceNon-Drinkable Water**
  + **A black screen with white text

    Description automatically generated with low confidenceDrinkable Water**

**หมายเหตุ :** เนื่องจากข้อมูลมีจำนวนมากจึงไม่สามารถแสดงออกมาทั้งหมดได้ เพราะจะทำให้เกิดความไม่สะอาดในการรายงานผู้จัดทำจึงกำหนดความกว้างของ leaf ไว้เพียง 80 หน่วยเท่านั้น

**Chart, scatter chart

Description automatically generatedChart, scatter chart

Description automatically generatedScatter Plot :**

ตัวแปรต้น : pH

ตัวแปรตาม : Turbidity

**Boxplot :**

**Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated Before remove outlier :**

outlier

**Chart, box and whisker chart

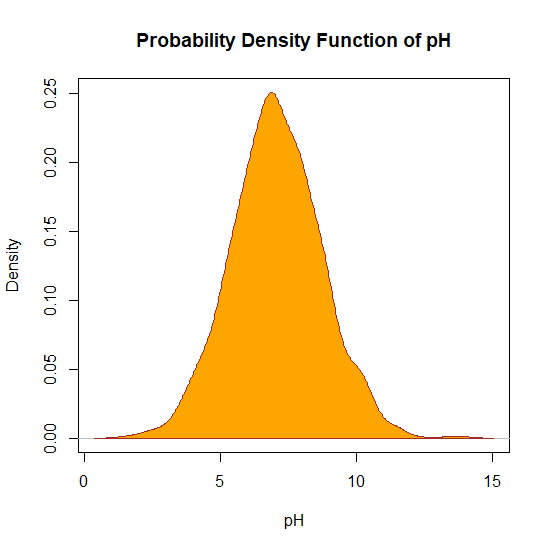
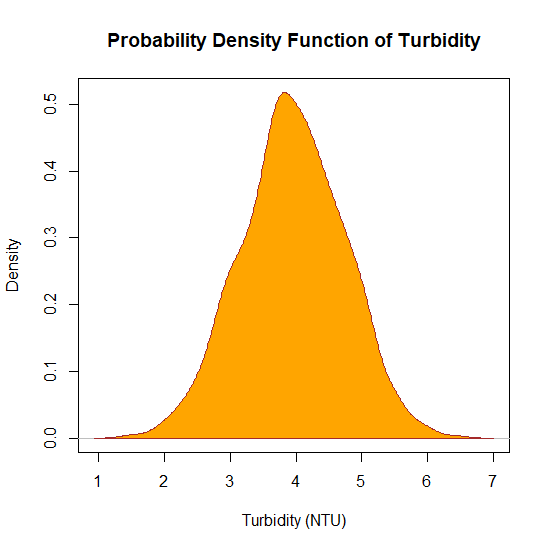
Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated After remove outlier :**

**Probability Density Function :**

** Non-Drinkable Water**

**Chart, histogram

Description automatically generatedChart, histogram

Description automatically generated Drinkable Water**

**Cumulative Distribution Function :**

**Chart, histogram

Description automatically generated Non-Drinkable Water**

**Chart, histogram

Description automatically generated**

**Chart, line chart, histogram

Description automatically generatedChart

Description automatically generated Drinkable Water**

**บทวิเคราะห์**

จากการวิเคราะห์ Histogram ของน้ำที่สามารถนำมาบริโภคได้ pH จะอยู่ในช่วง 5-9 ส่วนของน้ำที่บริโภคไม่ได้นั้นจะอยู่ในช่วง 4-10 จะเห็นว่าน้ำที่นำมาบริโภคได้นั้นโดยมากแล้วมีค่าเป็นกลางส่วนน้ำที่บริโภคไม่ได้มีค่าคววามกว้างของ pH มากกว่าคือ เป็นกรดแก่และเบสแก่ ซึ่งไม่สามารถนำมาบริโภคได้ โดยอ้างอิงจากน้ำที่นำมาบริโภคได้ต้องมีค่า pH ระหว่าง 6.5-8.5 ตามเกณฑ์ขององค์การอนามัยโลก

จากการวิเคราะห์ในส่วนของ Scatter Plot ของทั้งน้ำที่สามารถนำมาบริโภคได้และบริโภคไม่ได้ ค่าความขุ่น(Turbidity)จะอยู่ในช่วงเดียวกันแต่จะแตกต่างกันค่อนข้างชัดเจนคือค่าของ pH ที่ในส่วนของน้ำที่บริโภคได้นั้นจะเกาะกลุ่มในช่วง pH 6-8 และน้ำบริโภคไม่ได้จะเกาะกลุ่มในช่วง pH 4-8 อีกทั้งค่าของ pH ส่งผลต่อความขุ่นของน้ำโดย pH แปรผันตรงกับความขุ่น

จากการวิเคราะห์ในส่วนของ PDF และ CDF จะเป็นแนวโน้มเดียวกับ Scatter Plot คือน้ำที่สามารถบริโภคได้จะมี pH 6-9 และน้ำบริโภคไม่ได้จะเกาะกลุ่มในช่วง pH 4-10 และค่าความขุ่น(Turbidity) อยู่ใมนช่วง 3-5 NTU

**สรุปผล**

น้ำที่สามารถบริโภคได้นั้นจากกลุ่มตัวอย่างจะมีค่า pH เป็นกลางคือช่วง 6-8 และในส่วนของค่าความขุ่นจะอยู่ใน ช่วง 3-5 NTU โดยยังมีค่าที่ทับซ้อนกันอยู่บางส่วนซึ่งสามารถคำนวณเป็น Error ของการคำนวณได้แต่ก็ยากที่จะสรุปให้ลงตัวได้ ยกตัวอย่างเช่น ผลกระทบของ pH ต่อความขุ่นของน้ำ ทั้งนี้หากต้องการความชัดเจนที่มากขึ้นควรจะใช้คอลลัมน์อื่นๆเพื่อมาประกอบการพิจารณาด้วย เช่น จำนนวนของคาร์บอนที่ถูกสร้างจากแบคทีเรียในน้ำเป็นต้น

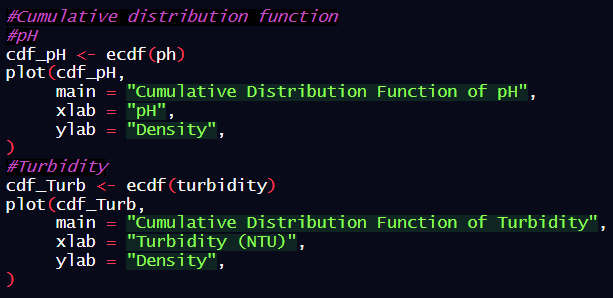
**Code :**

**A screenshot of a computer

Description automatically generated with medium confidenceText

Description automatically generatedA screenshot of a computer

Description automatically generated with medium confidence**

**Text, chat or text message

Description automatically generatedText

Description automatically generated with medium confidenceText

Description automatically generated**

**Source Code file : https://drive.google.com/file/d/182M48Lo26oLYOgRzpCp3cKMKcw1kes\_I/view?usp=sharing**