

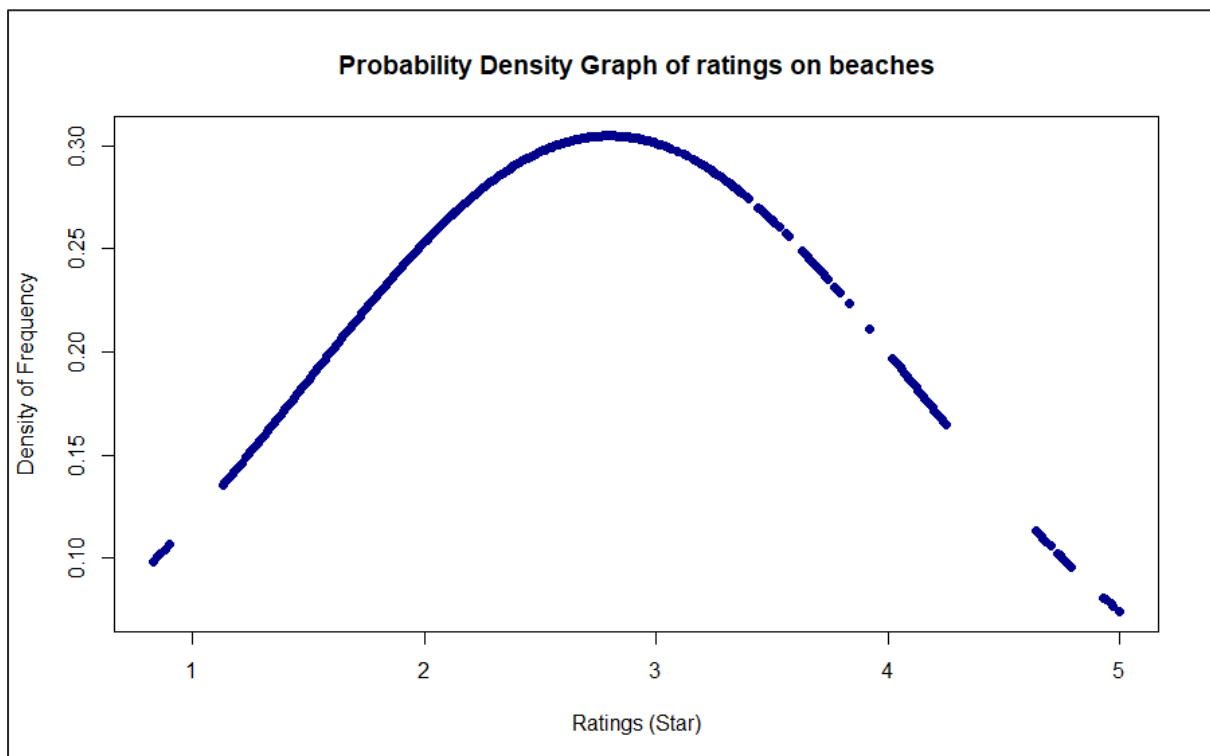
01076253 PROBABILITY AND STATISTICS HW 3

สืบเนื่องมาจาก HW 2 ที่ทางผู้จัดทำได้หยิบยกชุดข้อมูลเกี่ยวกับ Travel Review Rating Dataset (From the Machine Learning Repository of University of California) ซึ่งมี ข้อมูลการคำนวณข้อมูลเชิงสถิติพื้นฐานแต่ละคอลัมน์ ข้อมูลแสดงกราฟ Stem and leaf แผนภาพ Histogram Boxplot และ XY Scatter plot ของแต่ละคอลัมน์ ซึ่งทางผู้จัดทำได้เห็นข้อมูลที่เกิดจากการเรียงตัวของข้อมูลมากขึ้น ทำให้สามารถหาค่าสถิติพื้นฐานได้มากขึ้น ไม่ว่าจะเป็น ค่า Min Mean Max และ Quartile ของกราฟ จากข้อมูลเหล่านี้สามารถนำมาวิเคราะห์ เพื่อหาความสนใจและความต้องการของผู้ที่มีารีวิวสถานที่ของแต่ละประเภทได้ง่ายขึ้น และโดย HW 3 นี้ จะแสดงถึง Probability Density Graph และ Cumulative Distribution Graph ของแต่ละสถานที่เพื่อให้เห็นการกระจายตัวของข้อมูลให้ชัดเจนมากขึ้น

Average ratings on beaches

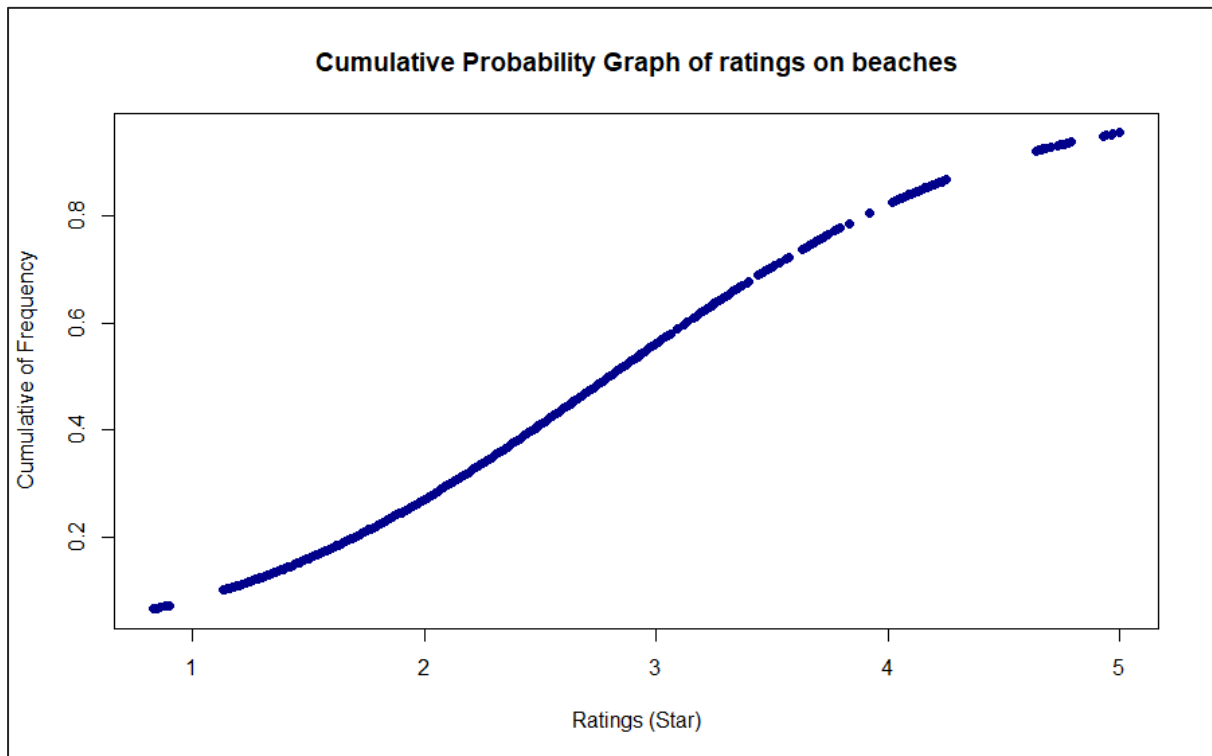
```
> mean(Google_Review_Ratings_$'Average ratings on beaches')
[1] 2.796886
> sd(Google_Review_Ratings_$'Average ratings on beaches')
[1] 1.309159
```

```
> DENS <- dnorm(Google_Review_Ratings_$'Average ratings on beaches', mean=2.796886, sd=1.309159)
> plot(Google_Review_Ratings_$'Average ratings on beaches', DENS, type = "p", col="dark blue", pch=19, main = "Probability Density Graph of ratings on beaches", xlab="Ratings (Star)", ylab="Density of Frequency")
```



Probability Density Graph of ratings on beaches

```
> CUMS <- pnorm(Google_Review_Ratings_$'Average ratings on beaches', mean=2.796886, sd=1.309159)
> plot(Google_Review_Ratings_$'Average ratings on beaches', CUMS, type = "p", col="dark blue", pch=19, main = "Cumulative Probability Graph of ratings on beaches", xlab="Ratings (Star)", ylab="Cumulative of Frequency")
```



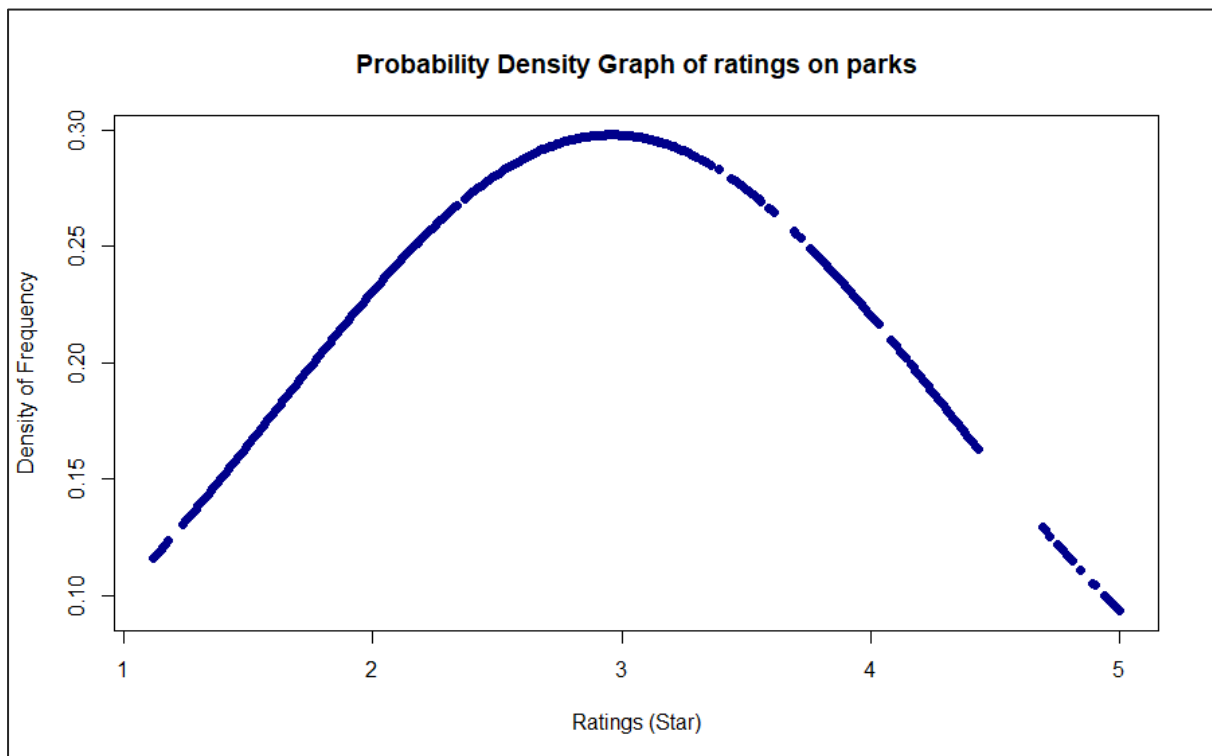
Cumulative Distribution Graph of ratings on beaches

1. หาค่า mean และ sd ของคอลัมน์ที่ต้องการจะทำการหา Probability Density
2. หลังจากนั้นใช้ตัวแปร "DENS" รับค่าคำสั่ง dnorm ซึ่งเป็นคำสั่งในการหา Probability Density โดยต้องใส่ค่า ข้อมูลของคอลัมน์ mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
3. จะได้กราฟของ Probability Density Graph of ratings on beaches มีหน่วยเป็น Star
4. ใช้ตัวแปร "CUM" รับค่าคำสั่ง pnorm ซึ่งเป็นคำสั่งในการหา Cumulative Distribution โดยต้องใส่ค่า ข้อมูลของคอลัมน์, mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
5. จะได้กราฟของ Cumulative Distribution Graph of ratings on beaches มีหน่วยเป็น Star

Average ratings on parks

```
> mean(Google_Review_Ratings_$'Average ratings on parks')
[1] 2.958941
> sd(Google_Review_Ratings_$'Average ratings on parks')
[1] 1.339056
```

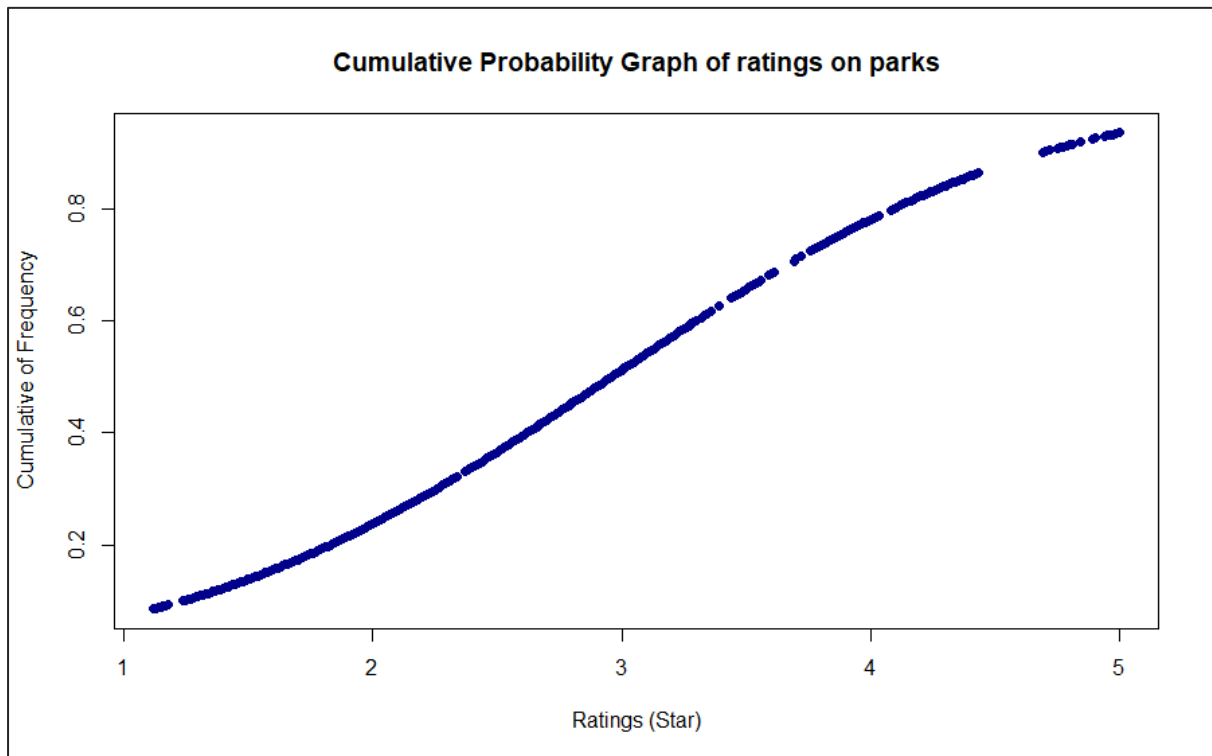
```
> DENS <- dnorm(Google_Review_Ratings_$'Average ratings on parks', mean=2.958941, sd=1.339056)
> plot(Google_Review_Ratings_$'Average ratings on parks', DENS, type = "p", col="dark blue", pch=19, main = "Probability Density Graph of ratings on parks", xlab="Ratings (Star)", ylab="Density of Frequency")
```



Probability Density Graph of ratings on parks

1. หาค่า mean และ sd ของคอลัมน์ที่ต้องการจะทำการหา Probability Density
2. หลังจากนั้นใช้ตัวแปร "DENS" รับค่าคำสั่ง dnorm ซึ่งเป็นคำสั่งในการหา Probability Density โดยต้องใส่ค่า ข้อมูลของคอลัมน์ mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
3. จะได้กราฟของ Probability Density Graph of ratings on parks มีหน่วยเป็น Star

```
> CUMS <- pnorm(Google_Review_Ratings_$'Average ratings on parks', mean=2.958941, sd=1.339056)
> plot(Google_Review_Ratings_$'Average ratings on parks',CUMS, type = "p",col="dark blue",pch=19 , main = "Cumulative Probability Graph of ratings on parks", xlab="Ratings (Star)", ylab="Cumulative of Frequency")
```



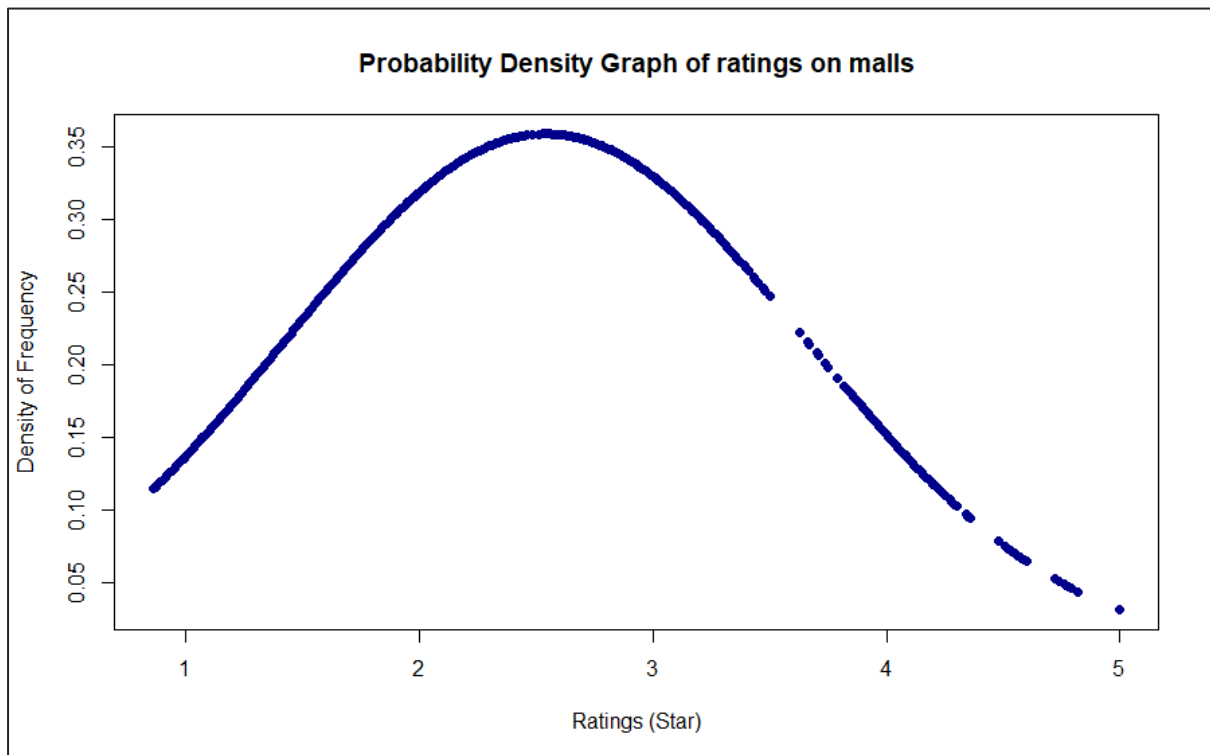
Cumulative Distribution Graph of ratings on parks

1. ใช้ตัวแปร “CUM” รับค่าคำสั่ง pnorm ซึ่งเป็นคำสั่งในการหา Cumulative Distribution โดยต้องใส่ค่า ข้อมูลของคอลัมน์, mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
2. จะได้กราฟของ Cumulative Distribution Graph of ratings on beaches มีหน่วยเป็น Star

Average ratings on malls

```
> mean(Google_Review_Ratings_$'Average ratings on malls')
[1] 2.540795
> sd(Google_Review_Ratings_$'Average ratings on malls')
[1] 1.111391
```

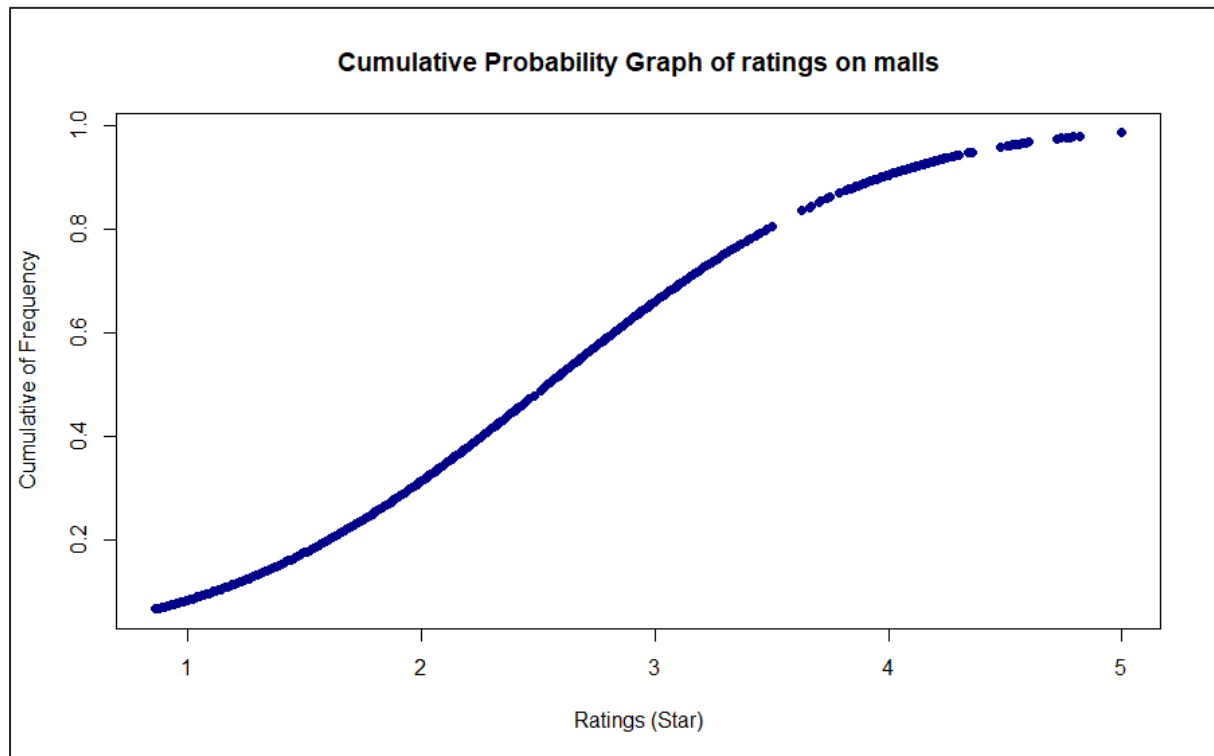
```
> DENS <- dnorm(Google_Review_Ratings_$'Average ratings on malls', mean=2.540795, sd=1.111391)
> plot(Google_Review_Ratings_$'Average ratings on malls', DENS, type = "p", col="dark blue", pch=19, main = "Probability Density Graph of ratings on malls", xlab="Ratings (Star)", ylab="Density of Frequency")
```



Probability Density Graph of ratings on malls

1. หาค่า mean และ sd ของคอลัมน์ที่ต้องการจะทำการหา Probability Density
2. หลังจากนั้นใช้ตัวแปร "DENS" รับค่าคำสั่ง dnorm ซึ่งเป็นคำสั่งในการหา Probability Density โดยต้องใส่ค่า ข้อมูลของคอลัมน์ mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
3. จะได้กราฟของ Probability Density Graph of ratings on malls มีหน่วยเป็น Star

```
> CUMS <- pnorm(Google_Review_Ratings_$'Average ratings on malls', mean=2.540795, sd=1.111391)
> plot(Google_Review_Ratings_$'Average ratings on malls',CUMS, type = "p",col="dark blue",pch=19 , main = "Cumulative Probability Graph of ratings on malls", xlab="Ratings (Star)", ylab="Cumulative of Frequency")
```



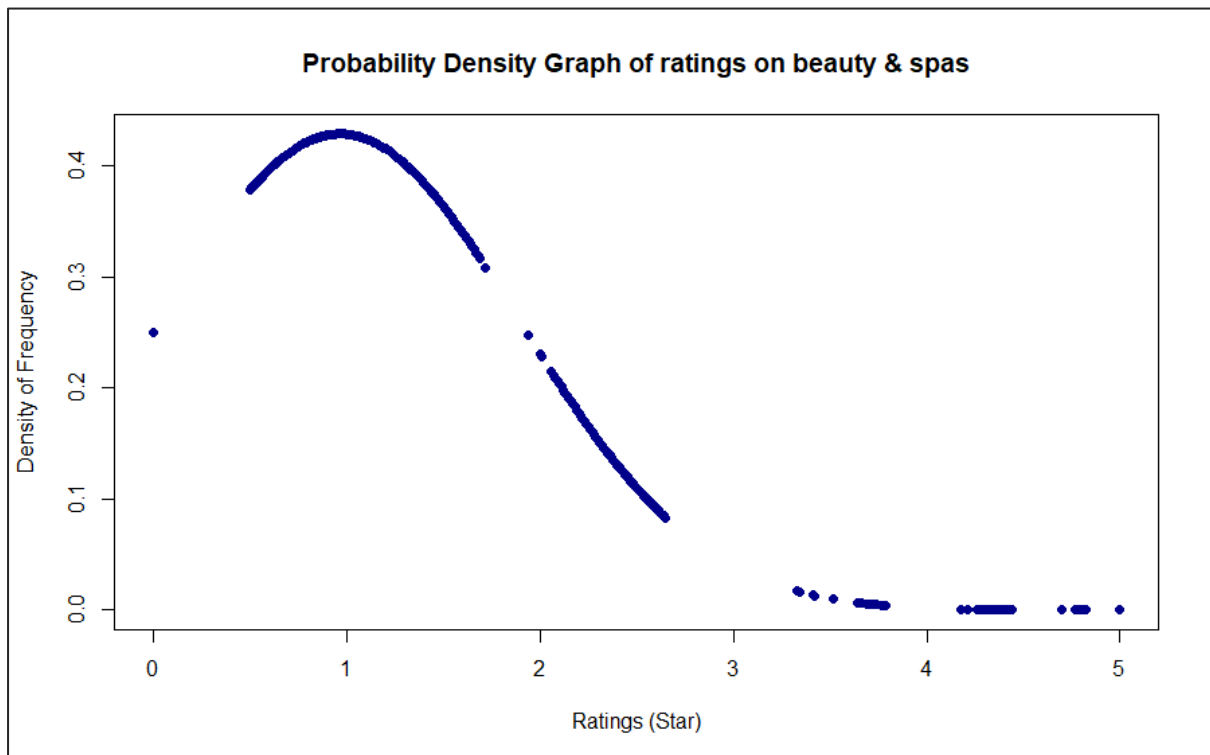
Cumulative Distribution Graph of ratings on malls

1. ใช้ตัวแปร “CUM” รับค่าคำสั่ง pnorm ซึ่งเป็นคำสั่งในการหา Cumulative Distribution โดยต้องใส่ค่า ข้อมูลของคอลัมน์, mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
2. จะได้กราฟของ Cumulative Distribution Graph of ratings on malls มีหน่วยเป็น Star

Average ratings on beauty & spas

```
> mean(Google_Review_Ratings_$'Average ratings on beauty & spas')
[1] 0.9658376
> sd(Google_Review_Ratings_$'Average ratings on beauty & spas')
[1] 0.9298533
```

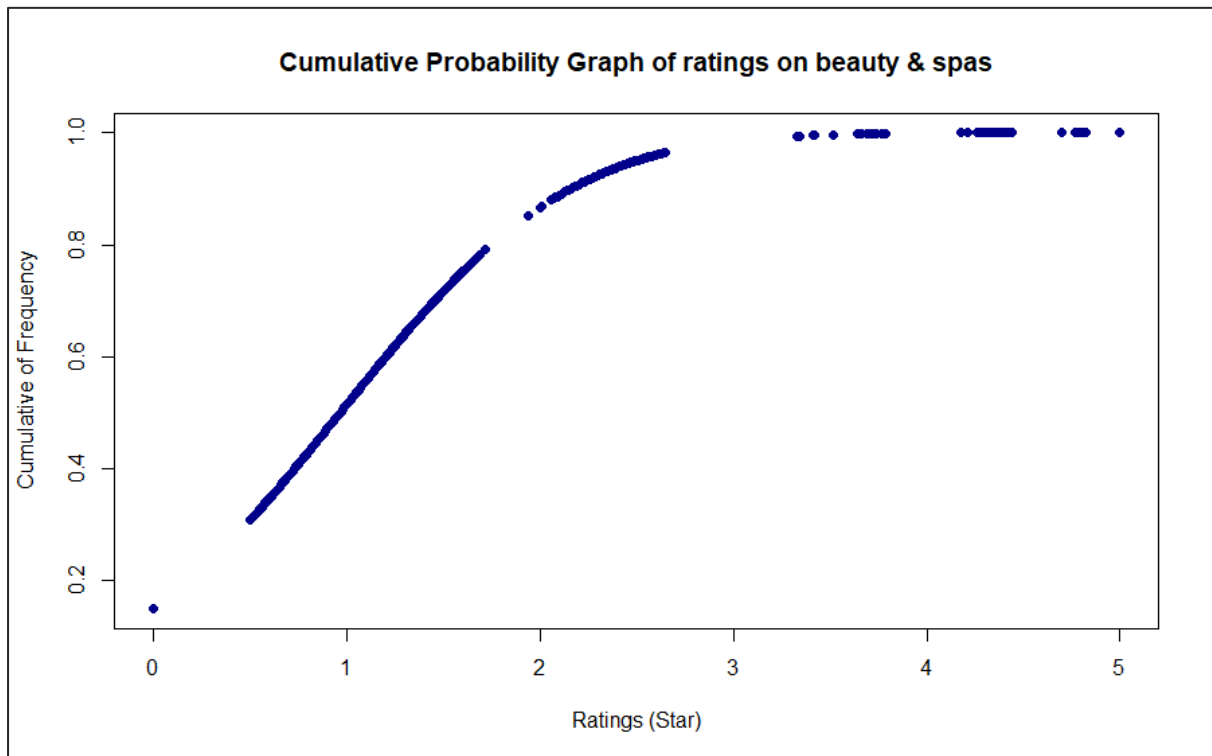
```
> DENS <- dnorm(Google_Review_Ratings_$'Average ratings on beauty & spas', mean= 0.9658376, sd=0.9298533)
> plot(Google_Review_Ratings_$'Average ratings on beauty & spas', DENS, type = "p", col="dark blue", pch=19, main = "Probability Density Graph of ratings on beauty & spas", xlab="Ratings (Star)", ylab="Density of Frequency")
```



Probability Density Graph of ratings on beauty & spas

1. หาค่า mean และ sd ของคอลัมน์ที่ต้องการจะทำการหา Probability Density
2. หลังจากนั้นใช้ตัวแปร "DENS" รับค่าคำสั่ง dnorm ซึ่งเป็นคำสั่งในการหา Probability Density โดยต้องใส่ค่า ข้อมูลของคอลัมน์ mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
3. จะได้กราฟของ Probability Density Graph of ratings on beauty & spas มีหน่วยเป็น Star

```
> CUMS <- pnorm(Google_Review_Ratings_$'Average ratings on beauty & spas', mean= 0.9658376, sd=0.9298533)
> plot(Google_Review_Ratings_$'Average ratings on beauty & spas',CUMS, type = "p",col="dark blue",pch=19 , main = "Cumulative Probability Graph of ratings on beauty & spas", xlab="Ratings (Star)", ylab="Cumulative of Frequency")
```



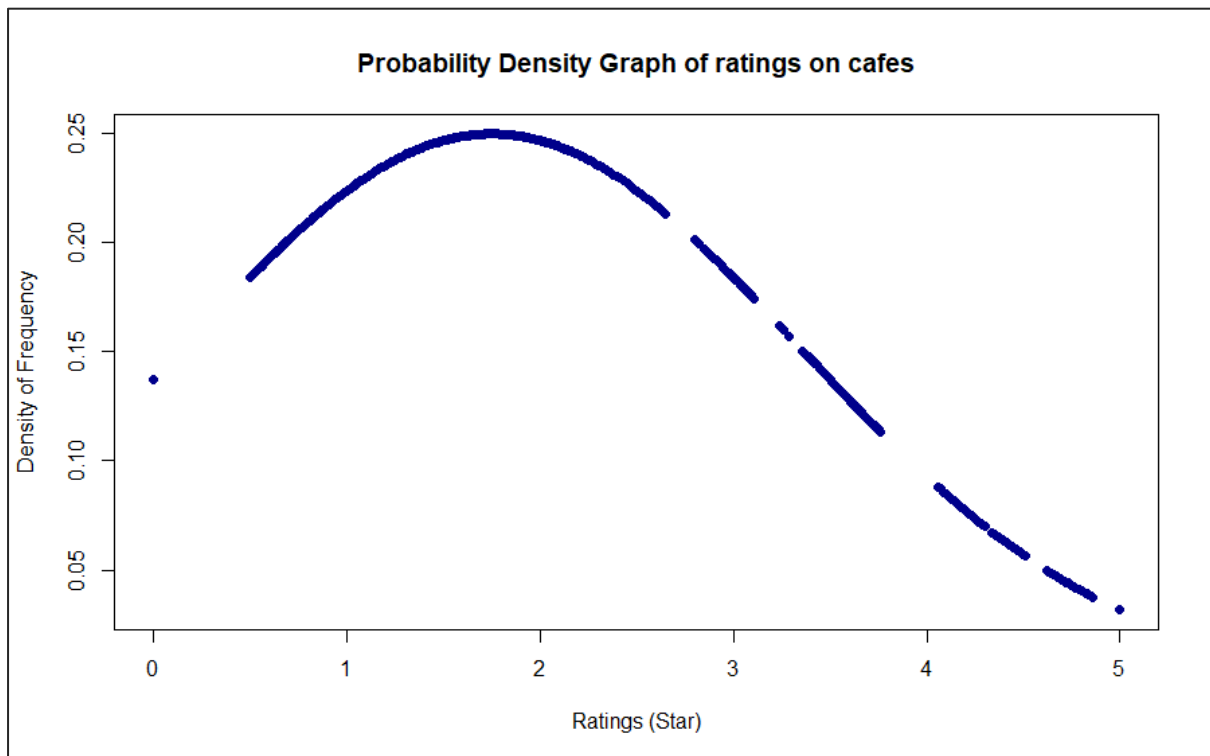
Cumulative Distribution Graph of ratings on beauty & spas

1. ใช้ตัวแปร “CUM” รับค่าคำสั่ง pnorm ซึ่งเป็นคำสั่งในการหา Cumulative Distribution โดยต้องใส่ค่า ข้อมูลของคอลัมน์, mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
2. จะได้กราฟของ Cumulative Distribution Graph of ratings on beauty & spas มีหน่วยเป็น Star

Average ratings on cafes

```
> mean(Google_Review_Ratings_$'Average ratings on cafes')
[1] 1.750537
> sd(Google_Review_Ratings_$'Average ratings on cafes')
[1] 1.598734
```

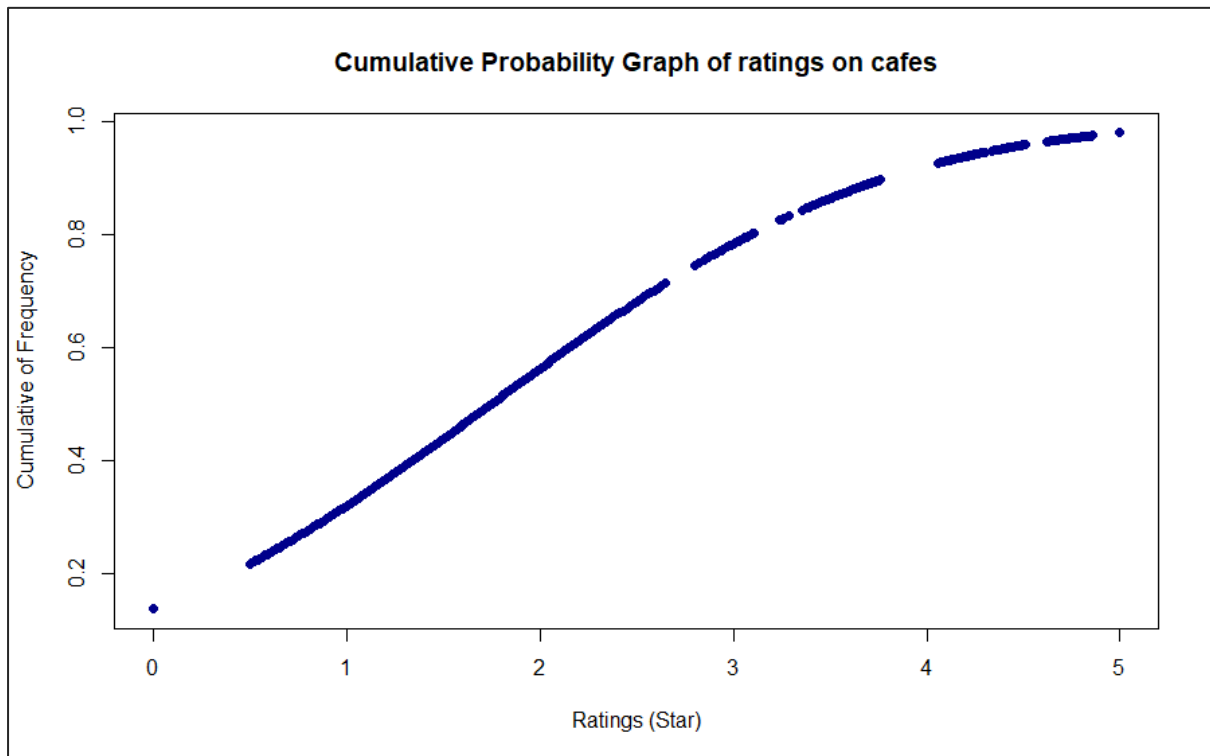
```
> DENS <- dnorm(Google_Review_Ratings_$'Average ratings on cafes', mean= 1.750537, sd= 1.598734)
> plot(Google_Review_Ratings_$'Average ratings on cafes', DENS, type = "p", col="dark blue", pch=19, main = "Probability Density Graph of ratings on cafes", xlab="Ratings (Star)", ylab="Density of Frequency")
```



Probability Density Graph of ratings on cafes

1. หาค่า mean และ sd ของคอลัมน์ที่ต้องการจะทำการหา Probability Density
2. หลังจากนั้นใช้ตัวแปร "DENS" รับค่าคำสั่ง dnorm ซึ่งเป็นคำสั่งในการหา Probability Density โดยต้องใส่ค่า ข้อมูลของคอลัมน์ mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
3. จะได้กราฟของ Probability Density Graph of ratings on cafes มีหน่วยเป็น Star

```
> CUM5 <- pnorm(Google_Review_Ratings_$'Average ratings on cafes', mean= 1.750537, sd= 1.598734)
> plot(Google_Review_Ratings_$'Average ratings on cafes',CUM5, type = "p",col="dark blue",pch=19 , main = "Cumulative Probability Graph of ratings on cafes", xlab="Ratings (star)", ylab="Cumulative of Frequency")
```



Cumulative Distribution Graph of ratings on cafes

1. ใช้ตัวแปร "CUM" รับค่าคำสั่ง pnorm ซึ่งเป็นคำสั่งในการหา Cumulative Distribution โดยต้องใส่ค่า ข้อมูลของคอลัมน์, mean และ sd และสุดท้ายทำการ plot() โดยค่าในแกน x เป็นคะแนนเต็ม 5 ดาว แกน y เป็นค่าการสะสมที่สอดคล้องกับค่า mean และ sd
2. จะได้กราฟของ Cumulative Distribution Graph of ratings on cafes มีหน่วยเป็น Star

บทวิเคราะห์ข้อมูลจากกราฟ

ชุดข้อมูลนี้มาจากที่เก็บแมชชีนเลิร์นนิงของมหาวิทยาลัยแคลิฟอร์เนีย, เออร์ไวน์ (UC Irvine) : ข้อมูลการจัดอันดับรีวิวการเดินทาง ชุดข้อมูลนี้จะถูกเติมโดยการจับคะแนนของผู้ใช้จากรีวิวของ Google รีวิวเกี่ยวกับสถานที่ท่องเที่ยวจาก 24 หมวดหมู่ทั่วยุโรปได้รับการพิจารณา ทางผู้จัดทำได้หิบบยคะแนนผู้ใช้ Google มีตั้งแต่ 1 ถึง 5 และมีการคำนวณคะแนนผู้ใช้เฉลี่ยต่อหมวดหมู่มากติวิเคราะห์

จากกราฟ Probability Density Graph ที่ทางผู้จัดทำได้แยกจัดทำแบ่งเป็น 5 ประเภทได้แก่ส่วนของ Average ratings on beaches, Average ratings on parks, Average ratings on malls, Average ratings on beauty & spas และ Average ratings on cafes วิเคราะห์ได้ว่า จากความหนาแน่นในแต่ละช่วงของคะแนนโดยรวมของทั้ง 5 ประเภท นั้นสามารถบ่งบอกได้ว่าจำนวน User ผู้รีวิวสถานที่แต่ละประเภท มีการให้คะแนนในแต่ละประเภทไปทางที่สูงมาก และมีคะแนนการรีวิวเฉลี่ยสูงเกินกว่าครึ่ง จาก User ผู้รีวิวสถานที่ทั้งหมดของ จากภาพโดยรวมจะเห็นว่าแนวโน้มในช่วง 0 – 3 ดาว มีความหนาแน่นสูง ทำให้เราสามารถวิเคราะห์ได้ว่า User ผู้รีวิวสถานที่ ส่วนใหญ่อยู่ในช่วง 0 – 3 ดาว ในประเภท Average ratings on beaches, Average ratings on parks, Average ratings on malls แต่ในประเภท Average ratings on beauty & spas และ Average ratings on cafes แนวโน้มในช่วง 0 – 1 และ 1 - 2 ดาว ตามลำดับ ซึ่ง User ผู้รีวิวสถานที่ การให้คะแนนมีความหนาแน่นสูงใน 2 ตามประเภทที่กล่าวมา

จากกราฟ Cumulative Distribution Graph ที่ทางผู้จัดทำได้แยกจัดทำแบ่งเป็น 5 ประเภทได้แก่ส่วนของ Average ratings on beaches, Average ratings on parks, Average ratings on malls, Average ratings on beauty & spas และ Average ratings on cafes วิเคราะห์ได้ว่าทั้ง 5 ประเภทมีความชันที่ไม่คงที่ และจะมีการกระจุยตัวของข้อมูลของช่วงต้นของกราฟมากกว่า กราฟในช่วงหลัง ซึ่งแสดงถึงการสะสมของ คะแนนเฉลี่ยของ User ผู้รีวิวสถานที่ ข้อมูลที่ได้จะสอดคล้องกับกราฟ Probability Density Graph ที่ได้ ยิ่ง Cumulative Distribution Graph มีความชันน้อย กราฟจะมีการกระจุยตัวของข้อมูลก็ยังมีค่าใกล้เคียง หรืออาจจะเท่ากัน อยู่มาก ถ้าความชันมาก และค่าต่างกันมาก ถึงเล็กน้อย กราฟจะมีความชันมากแบบต่อเนื่อง

สามารถสรุปได้ว่าการใช้กราฟ Probability Density และ Cumulative Distribution สามารถนำมาใช้วิเคราะห์ ข้อมูลจากแนวโน้มความหนาแน่นหรือการสะสมของข้อมูลที่สนใจได้