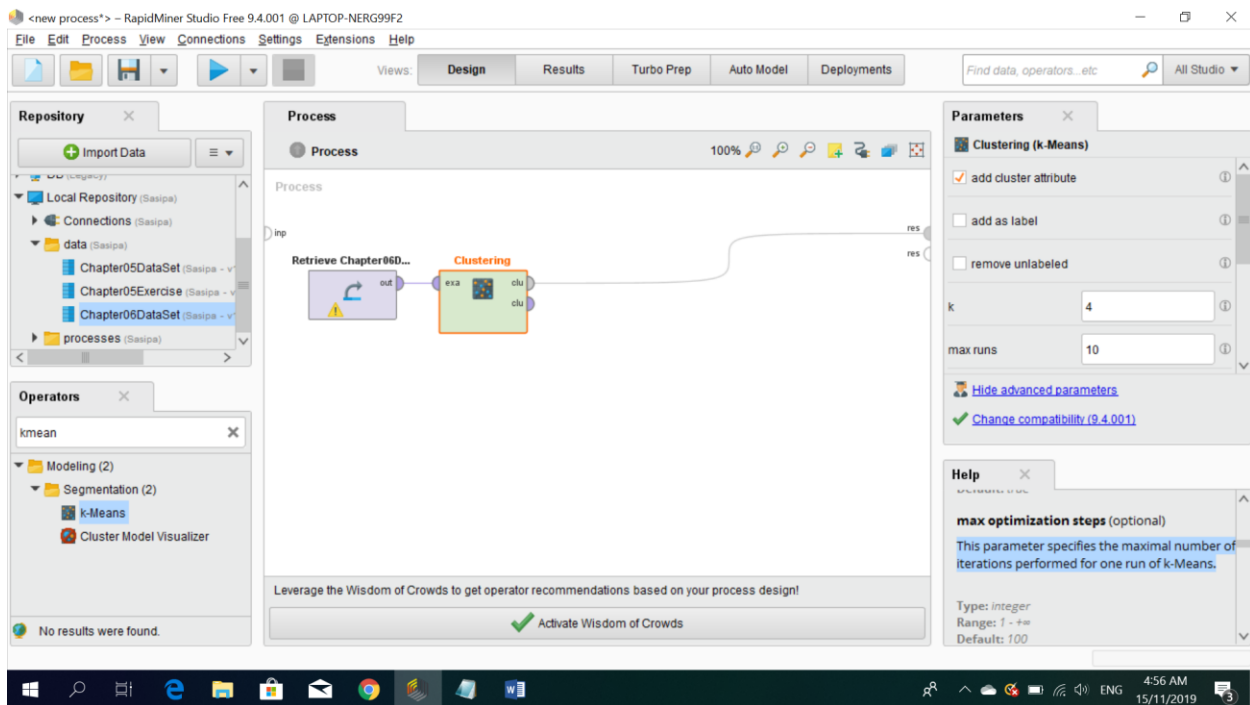


## Clustering

### K-means

Dataset - Chapter06Dataset.csv

- Cholesterol
- Weight (in pounds)
- Gender – 0 : female, 1 : male



บริษัทประกันต้องการดูข้อมูลลูกค้าเพื่อเฝ้าระวังเรื่องโรคหลอดเลือดหัวใจตีบ

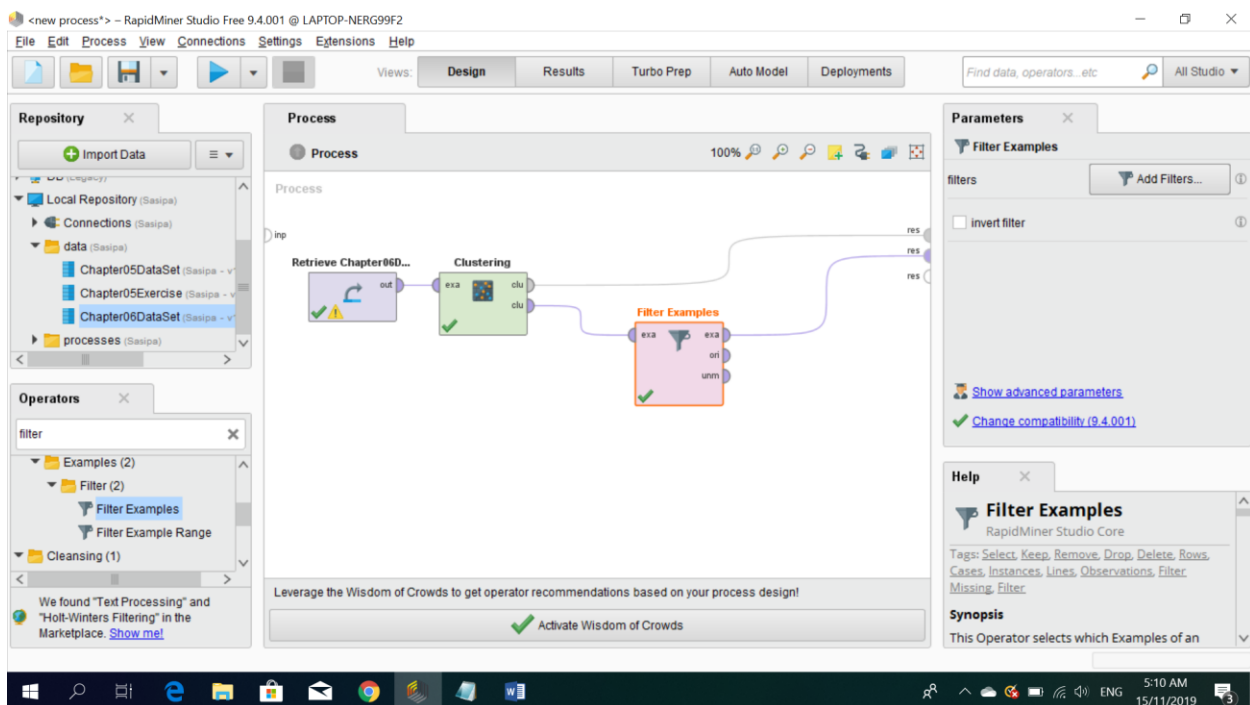
ต้องการดูลักษณะข้อมูลเพื่อหาว่ากลุ่มไหนที่เสี่ยงต่อการเป็นหลอดเลือดหัวใจตีบ

ทดลองสร้าง 4 cluster ลองดูผลลัพธ์ เลือกดู Centroid table

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
Weight	152.093	106.850	184.318	127.726
Cholesterol	185.907	119.536	218.916	154.385
Gender	0.441	0.543	0.591	0.459

ดูข้อมูลแต่ละคนได้ใน Folder View

เพื่อให้ดูข้อมูลคนในแต่ละ cluster ได้ทั้งกลุ่ม ให้เพิ่ม Filter Examples operator โดย add filters เลือก field และ value ที่ต้องการ



ดูผลลัพธ์ที่กรองได้ แล้วกดดู statistics สามารถนำช่วงข้อมูลของ cholesterol และ weight กลับไปค้นหาข้อมูลใน database ขององค์กรเพื่อดูข้อมูลของคน ๆ นี้ให้ละเอียดขึ้นได้

ใช้ dataset ชุดเดิม ลองเลือก k-medoids operator แล้วดูผลลัพธ์เปรียบเทียบกับทางเลือกใช้ k-means

ใช้ dataset ชุดเดิม ลองเลือก dbscan operator แล้วดูผลลัพธ์เปรียบเทียบกับทางเลือกใช้ k-means

### แบบฝึกหัด ใช้ RapidMiner

Dataset - Chapter06Exercise.csv

ข้อมูลของนักเรียน

- Student\_ID
- First\_Name
- Last\_Name
- Gender
- Grade
- Tardies - มาสาย
- Absences - ขาดเรียน
- Suspension - ให้ออกจากการเรียน
- Expulsion - ไล่ออก
- Arrest - ถูกจับกุม

ลองคิดว่าอยากจัดกลุ่มอย่างไร ทำแล้วดูผลลัพธ์

### แบบฝึกหัด เขียน python k-means

Dataset - Chapter06Exercise.csv