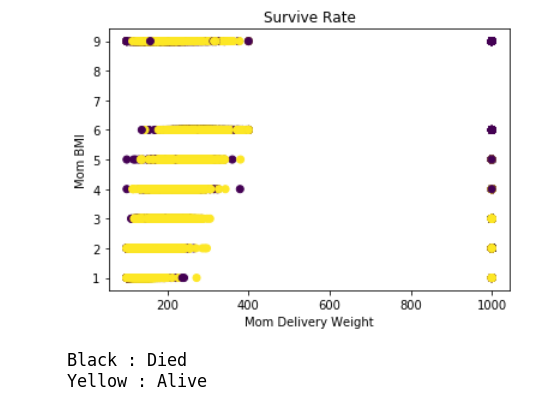
Big Data Qualification – JT – Justine Winata  
Data : Birth Data in 50 State

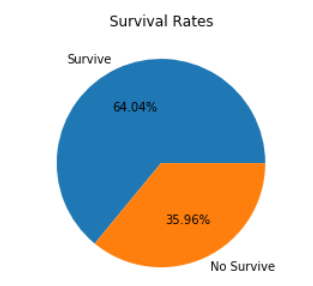
Predict Method : Clustering

Data to Select :   
- INFANT\_ALIVE\_AT\_REPORT (For data Testing)  
- INFANT\_SEX (Gender)  
- STEROIDS (Mother use steroids or no)  
- ANTIBIOTICS (Mother use antibiotics or no)  
- CIG\_BEFORE (Mother has ciggarate before)   
- MOTHER\_DELIVERY\_WEIGHT (Mother weight)  
- MOTHER\_BMI\_RECODE (Mother BMI)  
- APGAR\_10 (Inflant APGAR Score)  
- ANESTHESIA (Mother use anesthesia or no)  
  
All the selected data was concerned as the survival rates of the infant.

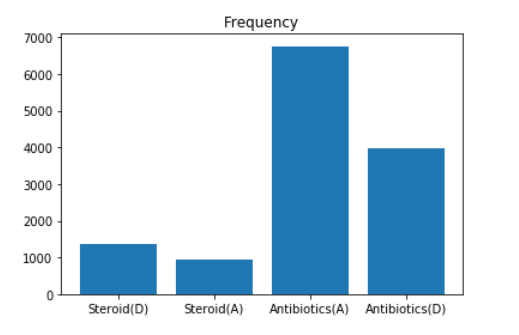
KMeans Setted to 2 (The cluster groups have 2 options rather that survive inflant or dead inflant)



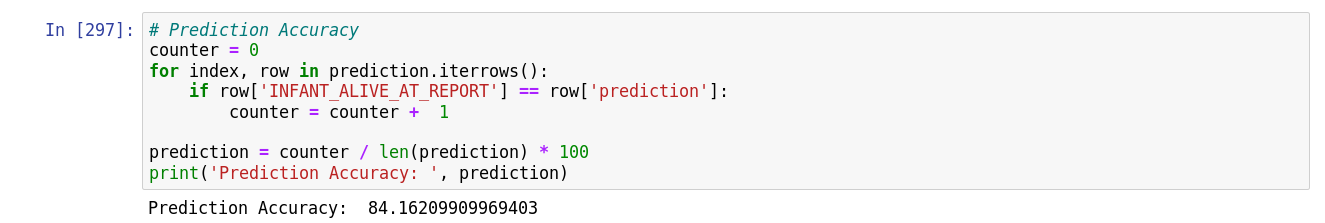
*Figure 1 Survive Rate With Mom BRI and Mom Delivery Weight*



*Figure 2 Prediction Survive Rate*



*Figure 3 Statistic Data of the death if Using Steroid and Antibiotics   
ex. Steroid(D) means 1000+ inflant was death because mother using Steroid*



*Figure 4 Predict Accuracy*

**Conclusion :**

Inflant survival rate will be increase 64.04% if the mother apply following term :

* Not using steroid
* Not using antibiotics
* Not using ciggerate
* Have a good weight
* Have a good BMI (Body Mass Index)
* Not using Anesthesia

Vice versa, if the mother didn’t apply the above term, then the inflant survival rate will decrease by 35.96%.