|  |  |  |
| --- | --- | --- |
| Understand data | Make first observation | 1 Passengerid is useless  2 Extract name info  3 Sex convert to numeric value  4 Missing age  5 Ticket is useful?  6 Carbin too many null, may drop this info  7 Embark has null  8 Embark convert to numeric value  9 Fare needs to be grouped  10 Age needs to be grouped |
|  | Use python to read basic data stats |  |
|  | Use tableau to visualize data from first observation | [https://public.tableau.com/profile/kenny2174#!/vizhome/Kaggle\_titanic\_15746524123480/agebin](https://public.tableau.com/profile/kenny2174" \l "!/vizhome/Kaggle_titanic_15746524123480/agebin) |
| Feature engineering plan | List to-do | 1 Remove passengerid  2 Extract name info (Mr. Miss.)  3 Sex convert to numeric value (1 male, 0 female)  4 Fill in missing age  5 Remove tickets  6 Remove Carbin  7 Fill in missing Embark  8 Embark convert to numeric value (S C Q)  9 Group fare (bin size )  10 Group age (bin size )  11 Fam\_size/sib/parch |
| Address issues in to-do list | Address issues in to-do list |  |
| Implement feature engineering |  | 1 Remove [PassengerId, Ticket, Cabin] |
|  |  |  |
|  |  |  |

<https://www.kaggle.com/startupsci/titanic-data-science-solutions>