|  |  |
| --- | --- |
| **1** | **Motivation & question** |
|  | I want to answer the Question:  Can you predict substance abuse and mental illness solely based on demographics?  Motivation:  Encourage more openness and discussion about taboo subjects. (I have minor depression and am worried if I mention it, it will affect job prospects for example.)  Interaction with the public |
| **2** | **Data collection** |
|  | Health information not publicly available in Australia.  Used US information – show webpage  Original dataset for 5 years 282,764 – 2,742  Community demographics may differ and may affect the outcome  Codebook – show codebook – 800 pages  Show MJEVER – options changed and removed  Mention error with marital status in survey – show WRKNUMJOB2 |
| **3** | **Data cleaning** |
|  | Show data summery page next to  ‘Datacleaning’ notebook in VS  Started models with 120,543 – 64 |
| **4** | **Machine learning models** |
|  | Open analytics page  Categorical data – used Logistic regression  Binary encoding – further encoding  Find a balance between questions and accuracy of prediction outcome – needed category importance  so used Random Forest  Several inputs – full info, partial info and just demographics  Added a grid search to all – finding the best settings for the model  Show the final categories – correlation diagram – 15 columns |

|  |  |
| --- | --- |
| **5** | **Machine learning outcomes** |
|  | Show the summery tables |
|  | Alcohol performed poorly - additional model – Sequential  One-hot encoding  Deep learning didn’t improve the outcome |
|  |  |
| **6** | **Conclusion** |
|  | Show dropdown options on survey page  Prediction with very limited amount of related data can fairly accurately predict substance use as long as it has grouped responses. |
| **7** | **Prologue** |
|  | Eventhough no prediction outcome plots  Show Introduction page  Plots on data to compare questionnaire outcome  Show Data page  Pie-graphs for questionnaire prediction input  US based – demographics may differ – review the input  Help page is not yet complete |