# Todo List

1. Task #1 R Shiny demo - Archana  -  60%, end of the week
2. ~~first demo of basis statistical functionality using Shiny and related R packages - target Monday~~
   * + 1. ~~Shiny walkthrough shiny.rstudio.com/tutorial   - 50%~~
       2. ~~Browse through shiny.rstudio.com/articles/basics.html    - 10 %~~
       3. ~~Watch all videos from this tutorial ( 3 parts) - 2 done, 40 %~~
       4. ~~Start building basic apps with Shiny and R - 50 %, Monday~~
3. Task #2 - Agile Scrum setup - Archana - product vision docs (Narayana), user story / break down of functionality (Narayana)- 10%, completion by next week
4. first demo of the tool and how we manage/track the tasks - 10%, Monday
   * + 1. Setup the tool, then Product Backlog, Stories, Team, Scrum etc - start after Monday
       2. Come up with a tentative workplan for  Product  #1 and #3 - - next week , Tue
       3. Make sure that all stake holders updated their tasks   - next week
5. Task #3 - Learning NLP - Yogesh  - 30%
   * + 1. ~~Go through~~ [~~NLP book~~](https://www.dropbox.com/home/rises.io%20Technical%20Team%20Folder/KnowledgeBase/Books?preview=NaturalLanguageProcessingWithPython.pdf) ~~- (Done)~~
       2. Show with NLP techniques /Demos - 30 %, 10 days
          1. ~~Occurrence stats (Done. Scrapes blogs, parses, prints frequency of a word)~~
          2. ~~Finding Similarity index between two texts (Done)~~
          3. ~~Reproduce blog with Python and MongoDB:~~ [~~https://www.analyticsvidhya.com/blog/2016/07/capstone-project/~~](https://www.analyticsvidhya.com/blog/2016/07/capstone-project/) ~~(Done)~~
          4. Uri incident and its influence on financial market(from Sept 28 till today)
          5. Analysing twitter data, predict cricket match outcome (Take 3rd match between India and NZ)
          6. Go through <http://www.datasciencecentral.com/profiles/blogs/5-text-classification-case-studies-using-scikit-learn>
          7. Go through <http://www.computerweekly.com/news/2240238397/Case-study-Hedge-fund-AHL-Man-Group-uses-MongoDB-to-feed-quants-with-data>
          8. Go through <https://www.analyticsvidhya.com/blog/2016/08/industry-insight-fighting-cyber-fraud-with-analytics/>
          9. Automatic summary of the 10 blogs,  even in crude form
          10. Word clustering
          11. Correlation and linking of topics
6. Task #4 DevOps   - Dev Env Setup 1. Windows10or7 2.Linux
   * + 1. Document how to setup the dev box for rises.io team, Open Src Analytics Platform
       2. [+Dev Env Setup](en/doc/W3yyIKVtzKiriSTMPW15K)
       3. Complete the setup at your own development machine
          1. R setup - complete before end of the week + add details to the doc - Yogesh
          2. Python setup - complete before end of the week + add details to the doc  - Archana -  90%
       4. Linux toolset for advanced analytics (if anything is only available on Linux) - Narayana
       5. Dev Env on Azure for Advanced Analytics  - Narayana
          1. Brief description about the technologies in HDInsight ecosystem and when to use what
          2. Simple regression problem walk-through  using HDInsight and Integrating with Microsoft ecosystem like Excel and Word
          3. Explore Cortana
7. Task #5: Documents Creation - for respective tasks/products
   * + 1. You can use Microsoft Office Online. Store it in DropBox. DropBox is linked to Microsoft Office online.
       2. Create in Open Office or Text working offline on local machine and then copy paste online.
       3. Need to categorise and refine (**standardization**)
8. a. Blog / Notebook / Research Paper (access = "**public**" )  - For specific formats for Blogs/Research Papers - Narayana
9. b. Presentation / Demo (access ="**restricted**" - for clients & investors)
10. For specific formats & expectations - Ajit
11. c. Requirements / Design Doc (access = "**internal**" - for team consumption) -
12. For specific formats - Ajit
13. Task #6: Blogs / Research Papers / Publications - Narayana
    * + 1. The Blog should be in editable document format, may be a paper on dropbox or doc on OneDrive
        2. It should be compatible with Wordpress to be published our website
        3. If you are planning use Jupyter, get some online help
        4. The blog will be posted on <http://rises.io/blog/>. This will be made public only after lot of reviews and appropriate refinements.
        5. All blogs should have images, illustrations, graphs wherever possible. There should be concluding comments, comparative numbers, etc. In short, all of our blogs should follow standard technical blog formats.
        6. From the technology baseline perspective all should follow our tech standardization - use of Anaconda, Microsoft R Open, Hadoop, Spark etc. Wherever possible make a Azure claim.

1. Product #1 - Fraud Detection - Archana  - 5%, started
2. first demo  - Archana to put realistic timelines after understanding what is expected, scope and target audience
   * + 1. Update, validate product vision - created, to validate
       2. Sprint planning - Agree on timelines & tasks
       3. Sprint execution & demo creation
       4. Reproduce the blog results from  <http://blog.revolutionanalytics.com/2016/09/fraud-detection.html>
       5. Integrate Shiny with the above developed code
       6. Implement Step 1 on Hadoop/Spark and note the computational time differences)
       7. Read through Chapter 4 : Detecting Fraudulent Transactions from the book (  <https://www.dropbox.com/home/rises.io%20Technical%20Team%20Folder/KnowledgeBase/Books?preview=Data+Mining+with+R_+Learning+with+Case+Studies.pdf> from page nos: 165 to 232) + Practice the R code mentioned in the chapter.
       8. Go through the blog: <https://www.mapr.com/blog/real-time-credit-card-fraud-detection-apache-spark-and-event-streaming>
       9. Go through the blog: <http://blog.revolutionanalytics.com/2016/02/data-science-deep-dive-with-revoscaler.html> - 20%
       10. Create a demo based on specs
       11. High level product definition (1page) - Narayana
       12. Analyse ‘outliers’ package in R  : <https://cran.r-project.org/web/packages/outliers/outliers.pdf> - 10%
       13. Go through <https://www.analyticsvidhya.com/blog/2016/08/industry-insight-fighting-cyber-fraud-with-analytics/>
       14. Presentation on Fraud Analytics  - Narayana
3. Product #2 - Intelligent Trader Assistant/ Robo Adviser - Narayana
4. first demo - ready to be published on web
   * + 1. Near time price prediction using real-time market data
       2. Learning trading patterns ML AI, suggesting profitable segments
       3. Study features from <http://www.turingfinance.com/dissecting-algorithmic-trading/>
5. Product #3.1 - Legal Tech AI - Yogesh
   * + 1. ~~Version 1: Read and strore acts in MongoDb~~
          1. ~~Indian Bare Acts:~~ [~~http://www.advocatekhoj.com/library/bareacts/~~](http://www.advocatekhoj.com/library/bareacts/) ~~are scraped by Narayana’s R script into csv, are imported into Mongo Db in structured way   -~~ [~~Data~~](https://www.dropbox.com/home/src/legaltech/data)
       2. ~~Version 2: Find Judgements (~~[~~Video~~](https://dl.dropboxusercontent.com/content_link/i4HT2hJgAcXTCNAL6kk4dsTzwZ9gMaKcDL8EPQtKPnGk4jdXc28jnTujAMUTq4FJ/file)~~)~~
          1. ~~Retrieve judgements matching the query (keywords, sections, etc) (Done)~~
          2. ~~Retrieve judgment similar to the given judgement based on text & citation similarity (Done)~~
       3. Version 3.1: Extract Summary - Yogesh
          1. ~~Extract Statistical summary by ready methods (Gensim, TextRank) (Done)~~
          2. Extract Legal domain specific Summary, new AI/DL based method
       4. Version 3.2: Predicting Case outcome - Narayana
          1. Build features and model to predict case outcome
6. Product # 3.2 - Legal Tech CRM - TBD
   * + 1. ~~Go through~~ <http://www.youwinlaw.com/> ~~and understand the functionality  (Done)~~
       2. Go through suiteCRM and install (Done, Installation notes [here](https://www.dropbox.com/home/rises.io%20Technical%20Team%20Folder/Trainings/YogeshSuiteCRM?preview=Notes_SuiteCRM.pdf))
       3. Change all the labels to rises name and call it "**court management suite"**
       4. Follow YouWinLaw for legal names for individual names
       5. Have three layer architecture: Data + Processing + UI
       6. Store Consumer Protect Act to MongoDB and access it from "**court management suite"** along with search feature (should be able to search Act from UI)
       7. Access Microsoft Office from "**court management suite",** i.e.,development documentation by access our inbuilt templates in MS Office
7. Service #1: SAS to R Migration
8. first demo - for client with same SAS code and mapping to R/Python/Open Src  - Narayana
9. first demo - for client with same SAS code and mapping to R/Python/Open Src  - Narayana
   * + 1. Source, Target and Transition state architecture, approach and business case  - Ajit
       2. Selection of code snippet & data to show case the capability - Narayana
       3. Solution accelerators, tools, R Analytics Lib - Narayana
10. Service #2: R/Open Src Adoption
    * + 1. Target state architecture, approach and business case  - Ajit
        2. Eco system demo - HDInsight (Hadoop, Spark) MRO, Anaconda Solutions on Azure
        3. Solution accelerators, tools, R Analytics Lib