Interactive data science for biotech

Oana C. Florean

Data Scientist

Nagarro

About me

- Graduated biochemical engineering
- Learned neurochemistry, Python and ML at Stockholm Uni
- Worked in Horizon 2020 research projects

Observed that R is preferred in

- Statistics and academic research- powerful for hypothesis testing, statistical analysis, publishing results
- Biotech and pharma- specialized frameworks for biological data, Bioconductor

R

- Developed in the '90s, inspired by S of the '70s at Bell Labs
- Interpreted and scripting language
- Wide variety of advanced statistical and graphical methods
- CRAN packages: dplyr- data manipulation, tidyr- data organization, caret- ML workflows, ggplot2- complex plots, plotly- interactive plots, tidyverse: suite for data science, shiny: interactive web apps
- 1D (vector, factor, list), 2D (matrix, data frame), and nD (array)
- Quirky syntax: assignment “<-”, pipe “%>%”, data structures access “$”
- Open-source, active community and plenty of resources

Shiny

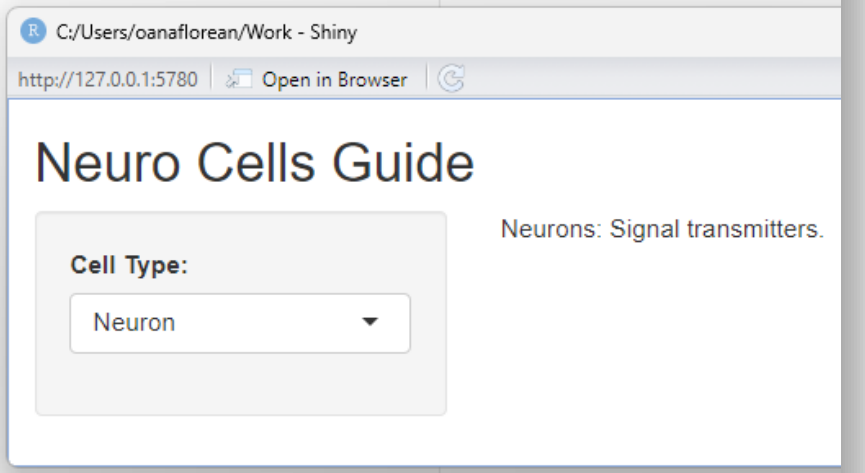
- Package to create interactive web apps
- Quickly and easily create dashboards directly from R
- Turn data into visualizations, analyses and products

Features

- Structured- the code is organized in separate UI and server sections
- Flexible- pre-built widgets or customized with HTML, CSS, JavaScript
- Reactive- the output automatically updates when the inputs change
- Accessible- hosted on a server and opened through any browser

Shiny

```r
# Load Shiny package
library(shiny)

# Define the UI
ui <- fluidPage(
  titlePanel("Neuro Cells Guide"),
  sidebarLayout(
    sidebarPanel(
      selectInput(inputId = "cellType",
                  label = "Cell Type:",
                  choices = c("Neuron",
                              "Astrocyte",
                              "Microglia"),
                  selected = "Neuron")
    ),
    mainPanel(
      textOutput("infoText")
    )
  )
)

# Define server logic
server <- function(input, output) {

  output$infoText <- renderText({

    descriptions <- list(
      Neuron = "Neurons: Signal transmitters.",
      Astrocyte = "Astrocytes: Neuron supporters.",
      Microglia = "Microglia: CNS immune defenders."
    )

    descriptions[[input$cellType]]
  })
}

# Run the app
shinyApp(ui = ui, server = server)
```



Interactive data science

Combines

- User interfaces- designed for clarity and insights
- Real-time exploration- adjust parameters and see results instantly
- Dynamic visualizations- charts that update with the data
- Collaborative work- multiple users, shared data

Exactly what biotech scientists need

- Complex data and analyses accessible to non-programmers
- Joint research accelerating discovery and innovation

The project

Requirements

- Platform for collaborative research
- Multiple analyses for data exploration
- Prototype in Shiny

Solution

- Comprehensive document app
- Versatile analysis modules

Dev team

- Software architect
- Data scientists- integrating statistics; iterative improvement
- Biostatisticians

Alzheimer's

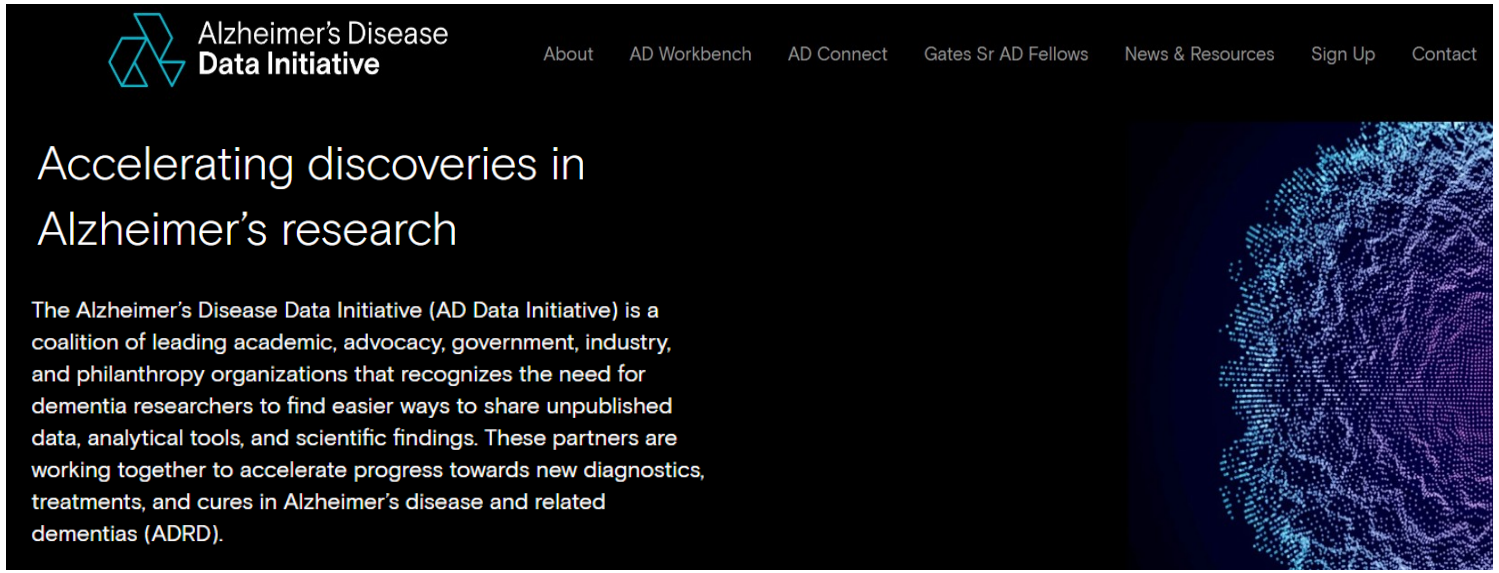
A neurodegenerative disorder leading to cognitive decline and dementia in older adults.

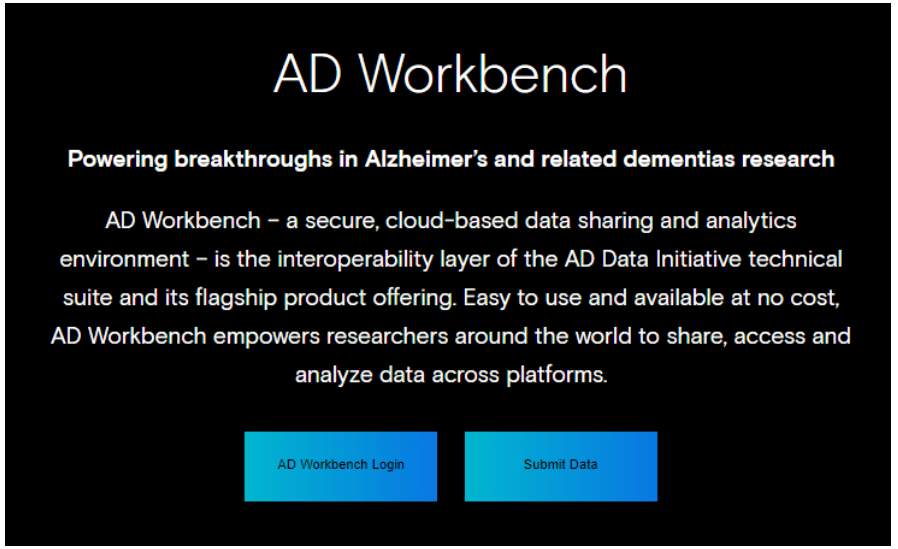
Biomarkers in cerebrospinal fluid

- Insights about disease progression and effects
- Neurogranin (synaptic plasticity), NfL (axon structure), α-Synuclein (neurotransmitters regulation), GFAP, S100B (astrocytes), IL-6, sTREM2 (inflammation)

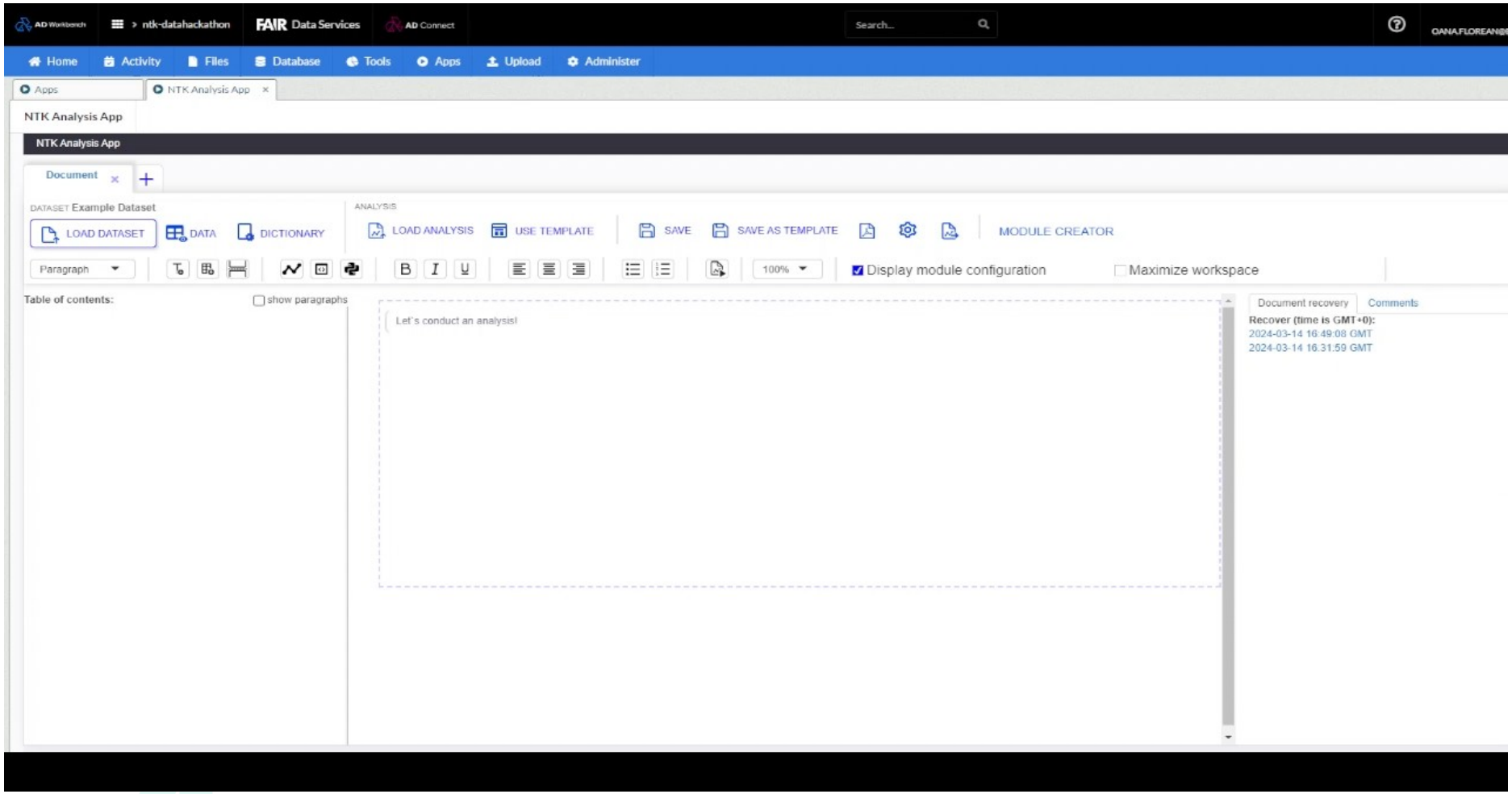
Alzheimer's research

https://www.alzheimersdata.org/

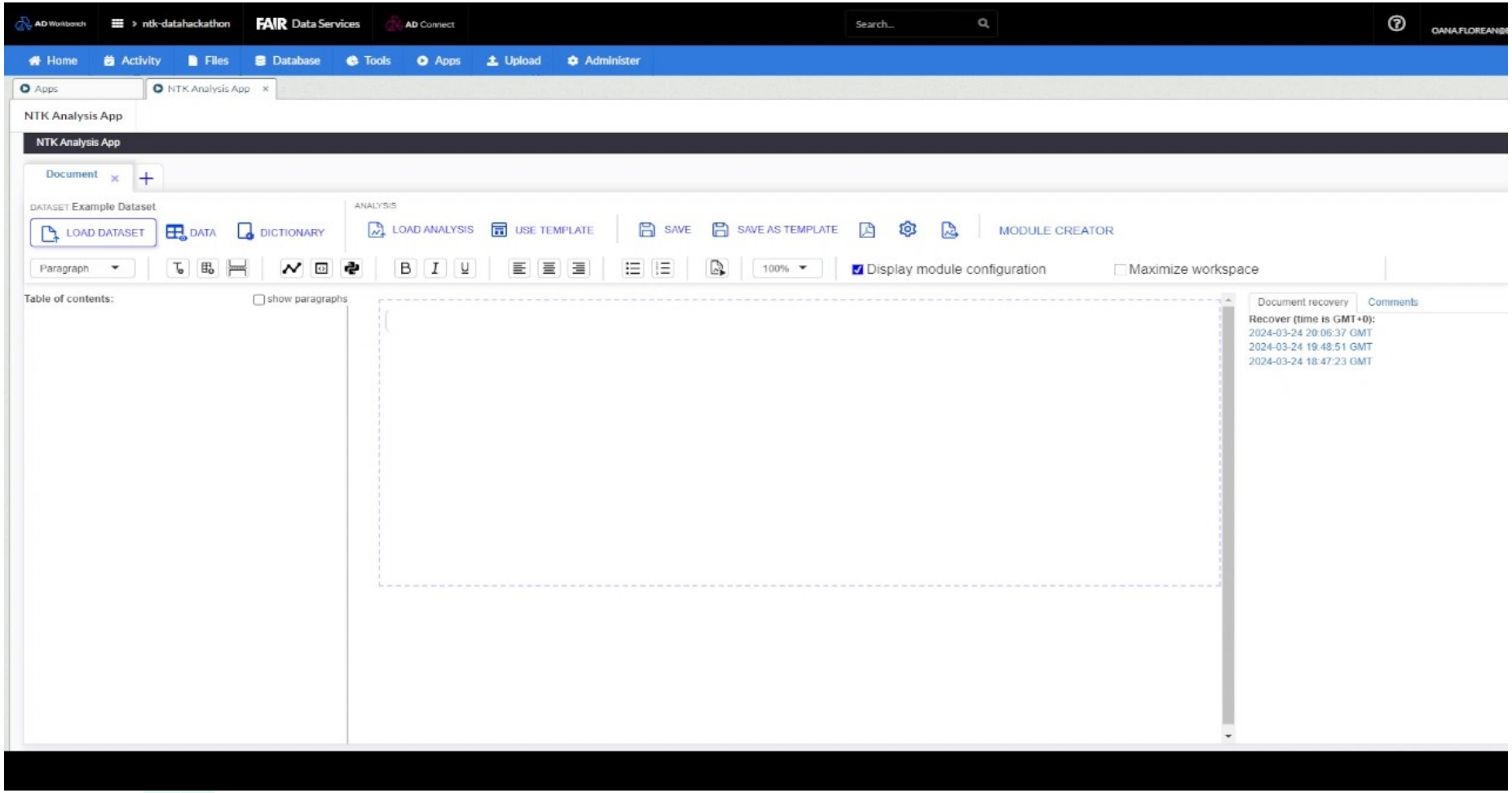




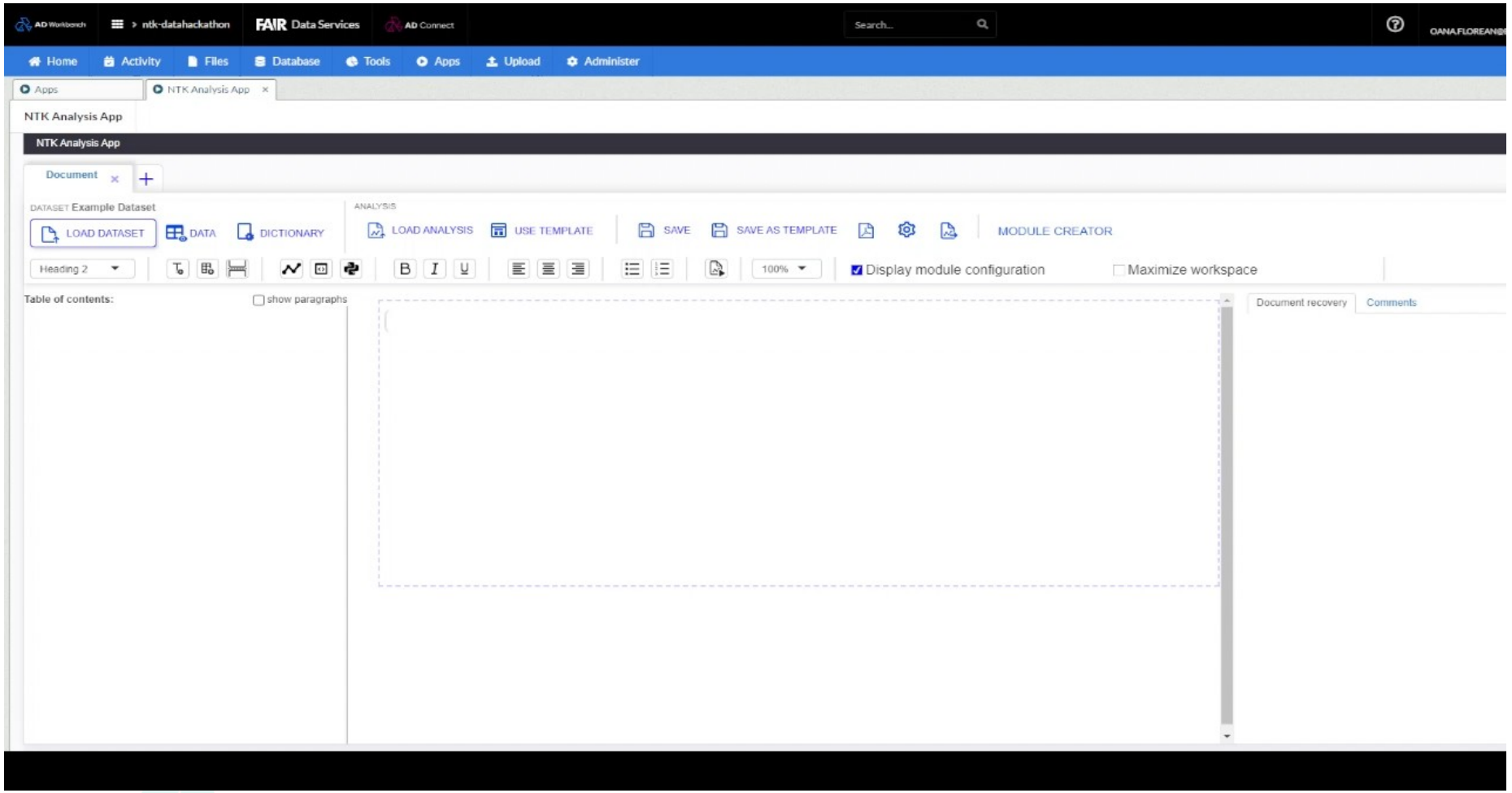
Analysis App



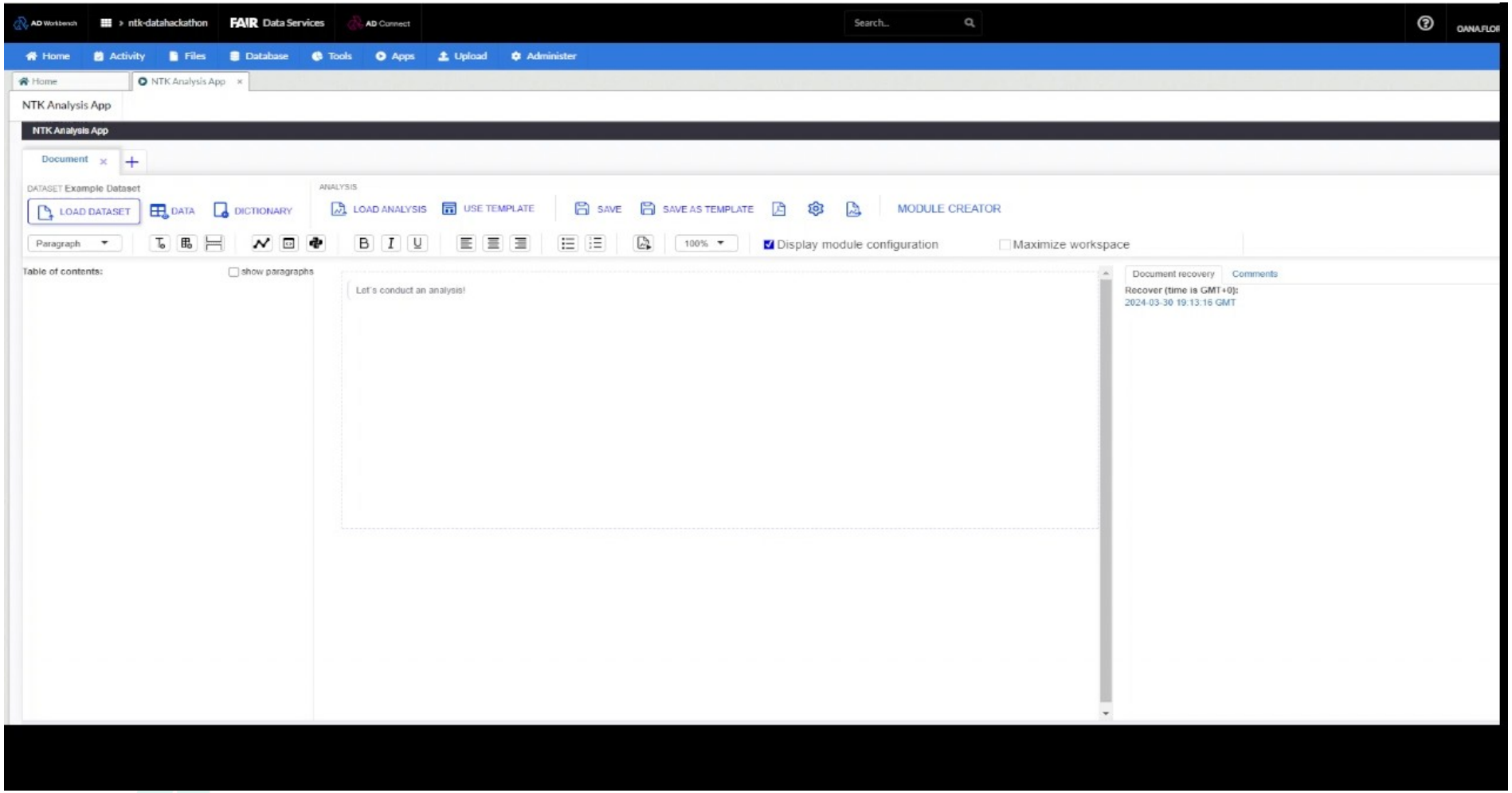
EDA module example 1



EDA module example 2



Hypothesis testing module



Achievements

- The report app with 40 modules for data analysis
- R package for extending Shiny functions
- A hackathon and a scientific publication
- Happy researchers and new collaborations

Interactive tools

- Streamlit- Python; quick dashboard prototype
- Dash- Python; production-ready dashboards
- Jupyter- tech users, education
- Voila- Jupyter notebooks for non-tech users
- Tableau, Power BI- business intelligence

Thank you!

What is the difference between ML and AI?

What is the difference between ML and AI?

ML is written in python/R

AI in powerpoint

☺