Library(tidyr)

long <-Wide%> % gather(Quarter, Revenue, Qtr.1:Qtr. 4) এখানে আমরা 'Gather' Function টি ব্যবহার করছি, যা tidyr এর Built-in-function। আরেকটা উদাহরণ দিই, আপনারা US Arrests Data Set টি ব্যবহার করে Practice করতে পারেন। my data <- USArrests [c(1, 10, 20, 30), ] my data এটা আমরা কাজের সুবিধার জন্য Data Set কে ছোট করে Subset এখন যদি আমরা একটা নতুন Column এতে যোগ করি সেক্ষেত্রে cbind Function টি ব্যবহার করতে পারি।

আরেকটি গুরুত্বপূর্ণ তথ্য : R এর কোন নতুন session গুরু করার আগে, আমরা সর্বপ্রথম একটি working directory তৈরি করবো। এটা একটা Folder যেখানে আপনি R session, data, installation directory টি set করা যাবে। Book Examples") করা যাবে । যেমন;

files গুলোকে Computer এর একটি নির্দিষ্ট Folder এ Save করে রাখবেন। ধরুন, আপনি যদি C drive এ users এর মধ্যে সব কিছু রাখতে চান, তাহলে setwd() ফাংশনটি ব্যবহার করে working setwd("C:/users/Munshi Naser/Documents/Data/R এখন এই C drive users এ আপনি যা রাখবেন, তা খুব সহজেই Rconsole এ কোন রকম File Path কে mention না করে Load My Data <- read.delim ("data file name.dat") অথবা, dataframe.name<-read.CSV ("filename.extention", header=TRUE) এখানে header = TRUE মানে হচ্ছে, প্রথম rowটি variable এর নাম নির্দেশ করে। খুব সহজ, তাই না! এছাড়া R-studio-তে file menu থেকে new project অপশনটির মাধ্যমে নতুন working directory তৈরি করা যায়। তারপর নতুন ওই directory drive folder এ আপনি আপনার ডাটা এবং Rscript এ save করে রাখতে পারবেন। সহজ বাংলায় R প্রোগ্রামিং 🌘 ৬৫

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৪৯

Article, Thesis এবং Book লিখার জন্য R- Mark Down ব্যাকআপ ফাইল থাকা জরুরি।

এই অধ্যায় শুরু করার আগে একটি গল্প সবার সঙ্গে শেয়ার করি সময়টা ২০০৭ সাল, তখন আমি বিশ্ববিদ্যালয়ের ছাত্র। আমাদের ্লাতক শেষ বর্ষের Thesis লেখার কাজ শুরু হয়েছে। কয়েকজন সহপাঠী মিলে কয়েকটি গ্রুপ করে গ্রুপ ভিত্তিক Thesis এর কাজ করছি। সারারাত জেগে থেকে Excel এ Data Input দিলাম, গ্রাফ তৈরি করলাম এবং এরপরে Microsoft word এ Copy করে গ্রাফণ্ডলো ঠিক করলাম। আমাদের কাজ শেষ করতে প্রায় সকাল হয়ে গেলো। আমরা সবাই কাজ শেষ করে সকালে ঘুমাতে গেলাম। বিশ্ববিদ্যালয়ের হলের কালচারটাই আসলে এমন। কোনো Assignment বা Thesis থাকলে শিক্ষার্থীরা রাত জেগে কাজ করে এবং দিনে ঘুমায়। যাইহোক, আমরা কাজ শেষ করতে পেরেছি বলে সবাই খুশি এবং ঠিক করলাম বিকেলে Thesis Paper প্রিন্ট দিতে যাবো। Thesis Paper নেওয়ার জন্য কম্পিউটারে পেনড্রাইভ প্রবেশ করালাম, অমননি ভাইরাস (কম্পিউটার ভাইরাস) অপারেটিং সিস্টেমকে ক্রাশ করে দিলো। তখন তো সবার মাথায় হাত! তারপর আবার অনেক কষ্ট করে সবকিছু করতে হলো। তখন মনে হলো সব ফাইলের একটা R প্রোগ্রামিং এমনই একটি প্লাটফর্ম যা শুধু আপনার Assignment বা Thesis এর কাজের ব্যাকআপই রাখবে না বরং Microsoft word বা Excel থেকেও কোনো কিছু কপি করা বা পেস্ট করার প্রয়োজন হবে না। R- Mark Down এমনই একটি প্যাকেজ যেখানে Microsoft Word এবং Excel এর প্রয়োজনীয় কাজ করা যায়। চিত্র-১২ এ R- Studio তে Mark Down File কিভাবে Open করতে হয় তা দেখানো হয়েছে।

এখানে আমরা Argument এ Data Frame এর নাম যা my data

key হচ্ছে Key কলাম এবং তার Value যা Output এ দেখতে চাই

এখানে State এর মানে সব State কে একসাথে এক Column

দেখতে চাই বা Gather করতে চাই। Value বলতে যেকোনো

Variable এর নাম, আর যদি অনেকগুলো Variable থাকে তাহলে

v1 : v10 দিতে হবে। আমরা যদি কোন Column কে বাদ দিতে চাই

তাহলে State add করে বাদ দিতে হবে। আমরা Output যা পাব

তা হচ্ছে, All column names (except state) Collapsed হয়ে

একটি Single key column "arrest\_attribute" এর Under এ

চলে এসেছে। এবং তাদের Value গুলো আরেকটি Column

Command ফলো করে একটু Practice করে ফেলুন!

৫০ 🌘 সহজ বাংলায় R প্রোগ্রামিং

৬৬ 
 সহজ বাংলায় R প্রোগ্রামিং

"arrest estimate" এর Under চলে গিয়েছে। এবার উপরের

এছাড়াও আপনি spread() Function দিয়ে একইভাবে Long ডাটা

Wider করতে পারবেন। separate() Function টি দিয়ে Single

column কে Multiple column এ Split করতে পারবেন। আবার

unite() Function দিয়ে Multiple column কে Single কলামে

আনতে পারবেন tidyr package টি দিয়ে। এই সবকিছু আবার একটি

করা হয়েছে। 

🕆 🔤 R- Mark Down প্যাকেজটি তৈরি হয়েছে মূলত R প্রোগ্রামিং ্যাংগুয়েজের টাইপ সেটিং প্লাটফর্ম হিসেবে। এই টাইপ সেটিংয়ে Microsoft Word এর মতো করে াইপ করা যাবে। যদি Table প্রয়োজন হয় তাহলে Table Insert রা যাবে, গ্রাফ প্রয়োজন হলে গ্রাফও তৈরি করা যাবে একই File এ রপর Report বা Assignment লখা শেষে PDF, Word এবং HTML Format এ পুরো Report Save করতে পারবেন। এভাবে চাইলে Python এ Jupyter Notebook ও আপনি Report বা Assignment তৈরি করতে পারবেন। চিত্র-১৩ এ R- Mark Down দেখতে কেমন হয় তা উল্লেখ

করা হয়। এটিও খুবই কার্যকর একটি Data Frame

এবার আরেকটি গুরুত্বপূর্ণ বিষয়, তা হল "missing values"। এই

missing values টি আমাদের দৈনন্দিন কাজে প্রায়ই Data Frame

এ আসে। এ বিষয়টি নিয়ে একটু কথা বলি। প্রথমে আমাদের যেকোনো

Data Frame এ Missing value আছে কি-না তা ভালভাবে Test

(চেক) করতে হবে। অল্প সংখ্যক ডাটার ক্ষেত্রে আমরা তা চোখে দেখে

বলে দিতে পারি। কিন্তু ধরুন, আপনার কাছে কয়েক হাজার, বা কয়েক

লক্ষ ডাটা আছে, তা থেকে Missing value চোখে দেখে বের করা

বেশ কঠিন। Data Frame থেকে Missing value বের করার জন্য

খুবই ছোট দুইটা Code শেয়ার করবো এবং কিভাবে Missing সংখ্যা

বাদ দিয়ে আমাদের Data Frame কে উপস্থাপন করে তা দেখাবো

খুব সহজ কমান্ড (is. na(x)) দিতে হবে। ডাটা ফ্রেম থেকে Missing

value এর সংখ্যা বের করার জন্য আমরা এই কমান্ডটি ব্যবহার করি

'na' টা সাধারণত Missing value নির্দেশ করে। আর কতগুলো

Missing value রয়েছে একটি ডাটা ফ্রেমে সেটা বের করতে

sum(is.na(df)) এই কমাভ ব্যবহার করতে হয়। আর যদি আমরা

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৫১

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৬৭

তো আমরা যদি Missing value এর Location খুঁজতে চাই, তাহলে

Transformation package

R- Mark Down Open করে প্রথমেই আপনি আপনার নাম, তারিখ, Title of the Report, Assignment, Thesis দিতে পারেন। his is the first R Book in Bangle. আমাদের Research Paper, Report এবং Thesis Paper লিখার প্রথমেই যেমন Introduction, Sub-Heading হিসেবে Research Objective লিখতে হয় তখন প্রথমে Heading যেমন Introduction লিখার শুরুতে # এই চিহ্ন দিয়ে শুরু করতে হবে। উদাহরণ # Introduction। একইভাবে Sub-Heading এর ক্ষেত্রে দুটি ## চিহ্ন এবং Sub-Sub-Heading এর ক্ষেত্রে তিনটি ### এই চিহ্ন বসবে। উদাহরণ: ##Research Objective এবং ### Research Problem। এভাবে একাধিক Heading, Sub-Heading আপনি দিতে পারবেন R- Mark Down এ।

৫২ 🌘 সহজ বাংলায় R প্রোগ্রামিং

৬৮ 🌘 সহজ বাংলায় R প্রোগ্রামিং

This is the first R Book in Research শব্দটি আমরা Bold করলাম। চিত্ৰ–১৬ \*italic\* \*\*bold\*\*

Report Writing কোনো শব্দ বা লাইন সহজে চোখে পড়ার জন্য বা শব্দকে আমরা যদি করতে চাই, সেক্ষেত্রে দিতে হবে। উদাহরণ: \*\*Research\*\*। এই কমান্ডটির অর্থ হলো

R- এ ব্যবহার করা খুব সহজ। R- যেহেতু Free; তাই প্রতিদিন R- এ

নতুন নতুন Function বা Package যোগ হচ্ছে এবং আমরা তা খুব

সহজেই শিখতে পারি Without paying anything...। আমার ইচ্ছা

আছে R- এবং Python এর একটি মডেল ভিত্তিক তুলনামূলক বই

লিখবো, যাতে খুব সহজেই ২টি ল্যাঙ্গুয়েজ এর মিল-অমিল খুঁজে

পাওয়া যায়। আর আমার ইউটিউব চ্যানেল তো রয়েছে আপনার

পাশে। আজকালের সবচেয়ে জনপ্রিয় দুইটি শব্দ হচ্ছে ডাটা সায়েন্স

্রএবং মেশিন লার্নিং। কিন্তু অনেকেই আমরা জানি না যে এট

পরিসংখ্যান আর গণিতের মডেল বিশ্লেষণ ছাড়া তেমন কঠিন কিছু না।

যেহেতু আমি অর্থনীতির ছাত্র, তাই আমি অনেক বুঝার চেষ্টা করেছি যে

এটা কি এমন কিছু যা কম্পিউটার সায়েন্স বা IT ডিগ্রি ছাড়া বুঝা যাবে

না। কিন্তু আমার ভুল ভাঙলো যখন আমি নিজে পড়তে শুরু করলাম।

একজন Non IT বা Non-Computer Science ব্যাকগ্রাউন্ডের মানুষ

হয়ে আমি কি বুঝলাম আসুন একটু শেয়ার করি; অর্থনীতির খুবই

গুরুত্বপূর্ণ বিষয় হচ্ছে ইকোনোমেট্রিক্স যা গণিত, অর্থনীতির কিছু মূল

Idea এবং পরিসংখ্যান নিয়ে তৈরি হয়েছে। এই বইয়ে আমি

Econometrics এর মূল কাজ এবং একটি Model সংক্ষেপে উল্লেখ

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৫৩

Down। যা দিয়ে আপনি খুব সহজেই আপনার Research Paper, Report এবং Thesis এর যথাস্থানে Table বা Graph তৈরি করতে চিত্ৰ-১৮ তে তা দেখানো হয়েছে। সহজ বাংলায় R প্রোগ্রামিং 🌘 ৬৯ ৭০ 🌘 সহজ বাংলায় R প্রোগ্রামিং

এছাড়াও আপনি যদি আপনার Paper এ লেখার মাঝে Table, Graph বা Data Entry করতে চান তাহলে Computer keyboard এ Control+Alt+I এই কমান্ডটি দিতে হবে। এই Command টির মাধ্যমে একটি Code chunk add করার সুযোগ দিবে R Mark া করে শুধুমাত্র করতে চান, তাহলে Code chunk এর Line  $\P$  {r,echo = alse} দিতে হবে।

হয় আপনার Guess করা একটু কঠিন বা ভুল হয়ে যাবে। তবে

আপনার কাছে যদি সব বা পূর্বের বেশিরভাগ IPL এর ম্যাচের তথ্য

থাকে তাহলে তা বিশ্লেষণ করে আপনি কিন্তু ভালো একটি Guess

করতে পারবেন। তো এই Guess টি ভালো কিনা তা বুঝবেন কি

করে। মনে করুন, আমরা চাচ্ছি, IPL এর কয়েকটি ম্যাচের Average

Six (ছক্কা) হবার সাথে IPL Tournament এর Average ছক্কার

আমাদের কয়েকটি ম্যাচের উপর ভিত্তি করে করা Average বা

পরিসংখ্যানের ভাষায় Mean টি একটি ভাল Estimation প্যারামিটার

হবে। মূলত কয়েকটি ম্যাচ যদি হয় আমাদের Sample, তাহলে

IPL পুরো Tournament) এর চরিত্র কতটা ভালোভাবে উপস্থাপন

করতে পারছে তা গুরুত্বপূর্ণ। অন্যদিকে Prediction বা অনুমান য

মেশিন লার্নিংয়ের একটি অন্যতম প্রধান কাজ, তা হচ্ছে আমরা IPI

এর বেশিরভাগ পূর্বের ম্যাচ দেখে বা পূর্বের Tournament পর্যালোচন

করে কতোগুলো Six (ছক্কা) হতে পারে ভবিষ্যতের ম্যাচ গুলোতে তা

অনুমান করা। এতে মূলত অতীত ডাটার উপর ভিত্তি করে ভবিষ্যতের

তথ্য অনুমান করা হয়ে থাকে। তো পরিসংখ্যানের Model এর উদ্দেশ্য

৫৪ 🌘 সহজ বাংলায় R প্রোগ্রামিং

প্যারামিটার Estimation এর কাজ হলো Population (এই ক্ষেত্রে

কাছাকাছি Guess করবে। যদি আমরা তা করে ফেলি, তাহলে

চিত্র ১৮ এর কমান্ডটি ব্যবহার করে PDF বা Word document এ এতে কোনো chunk Command Show করবে না এবং আউটপুট দেখা যাবে খুব সহজেই। ## speed dist ## Min. : 4.0 Min. : 2.00 ## 1st Qu.:12.0 1st Qu.: 26.00 ## Median :15.0 Median : 36.00 ## Mean :15.4 Mean : 42.98 ## 3rd Qu.:19.0 3rd Qu.: 56.00 ## Max. :25.0 Max. :120.00 এবার আসা যাক R- Mark Down এ Assignment, Report বা Thesis লিখতে কিভাবে Reference add করা যায় এই বিষয়ে। একটি ভালো Assignment, Report বা Thesis লিখতে অবশ্যই আপনাকে Reference দিতে হবে। কারণ Reference ছাড়া আপনার Research অসম্পূর্ণ থাকবে। প্রথমে আপনি Zotero, Endnote Mendeley reference manager software থেকে আপনার Reference গুলোর Bibtex File Download করবেন। এরপর R-Mark Down এ Bibtex File টি Upload করবেন। Step গুলো হলো : প্রথমেই আপনার R- Mark Down এর Output File এর নাম bookdown::html-document:default করে নিতে হবে। আপনি চাইলে Bookdown নাও করতে পারেন। Bookdown R-এর একটি Book writing package. Markdown বা Bookdown package দুটি প্রায়ই একই রকম এবং একই ধরনের Command লাইন Follow করে। আমার Youtube Channel এ

তারপর Bibtex File টি Same R- Working Directory তে

রাখবেন। যেখানে আপনি R- এর সব File Save করেছেন। এখন

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৭১

Network fewer assumption | Model with lots of assumpt

Learning from past to Model fitting to estimate

Generalization algorithm | Test a set of performance

dependent diagonistic tools

Supervised learning Regression/Classification

Uses mathematical equation Uses mathematical equation

bility to predict the outcome | Estimator seeks to know

of a random variable property of the true state of

আগেই বলেছি এই বইয়ে Machine Learning বা Econometrics

nature (i. e. population)

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৫৫

Unsupervised learning Density estimation/Clust

Strong prediction ability Best "Estimate"

predict future (Specially linear line)

চিত্ৰ-২০.১ output: bookdown::html\_document2: default এই বিষয়ে ভিডিও রয়েছে, চাইলে দেখে নিতে পারেন।

এই Code টি দিতে হবে; bibliography:Your-bib-file.bib সর্বশেষে intext citation দিতে হবে। অর্থাৎ Assignment Report বা Thesis এ লেখার ফাঁকে ফাঁকে Reference add করতে হবে। যেমন @anderson2015। এখানে @ চিহ্নের মানে Bibtex File এর anderson2015 এর Full Reference এর সাথে আপনার Text Reference এর সমন্বয় করা হলো। ব্যাস, R- Mark Down এ হয়ে গেল Reference writing এর কাজটি। খুবই সহজ তাই না! এবার নিচের চিত্র-২০ এ Step গুলো একসাথে দেখে নেই। # make sure bookdown is installed install.packages("bookdown")

Wooldridge এর Econometrics বইয়ের ডাটা এই ক্ষেত্রে ব্যবহার

data (ceosal1, package = 'wooldridge') attach (ceosal1)

এটাও একটা পদ্ধতি R Console এ Data frame কে ব্যবহার করার।

তো parameter estimate করার জন্য Value গুলো বের করি,

COV(roe, salary)

Var(roe)

mean(salary)

mean(roe)

উপরের Command এর Alternative, install packages

("Wooldridge") library (Wooldridge) data (ceosal1)

অথবা Regression এর জন্য R একটা বিশেষ ফাংশন তৈরি করেছে

যা lm নামে পরিচিত। lm ফাংশনটি দিয়ে খুব সহজে linear

view (ceosal1) করা যাবে]

regression রান করি,

৫৬ 🌒 সহজ বাংলায় R প্রোগ্রামিং

৭২ 🌘 সহজ বাংলায় R প্রোগ্রামিং

3. Add a Link to Your BibText File to the YAML Header Using the code below, add a link to your BibTex file. bibliography: your-bib-file.bib চিত্ৰ-২০.৪ Now your entire YAML header looks like this: title: "Your title here" author: "Your name here" bookdown::html\_document2: default bibliography: your-bib-file.bib

251duals: Min 1Q Median 3Q Max -1.6093 -0.7505 -0.0822 0.9351 1.3898

ব্যবহার করতে পারি। Code টি হবে;

2. Create a BibTex File Containing References

example of a BibText formatted citation is below.

Manual{anderson2015,

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 '

idual standard error: 1.116 on 8 degrees of freedom tiple R-squared: 0.9683, Adjusted R-squared: 0.964 tatistic: 244.3 on 1 and 8 DF, p-value: 2.8e-07

এবার যদি Plot করতে চাই, তাহলে আমরা abline() Function টি

with(ceosal1, plot (roe, salary, abline(myolsresult), ylim

যেহেতু অনেক বড় ডাটা সেট, তাই আমরা দেখার সুবিধার জন্য Y কে

ormatted, machine readable reference list. You can create BibTex reference lists in

manager tools including **Zotero** and **Mendeley** or you can go the hard route

= c(o, 4000)))

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৫৭

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৭৩

to be saved using the UTF-8 encoding. Finally you can add citations to a report. To do this you use the syntax @anderson2015 title = {Exhumation by debris flows in the 2013 Colorado Front Range storm}, author = { S.W. Anderson and S.P. Anderson and R.S Anderson}, এখন আপনার Assignment, Report বা Thesis লিখার পর F Mark Down এ Kint Option এ ক্লিক করলে আপনার Output File টি PDF, HTML কিংবা চাইলে Word এ Show করবে। এই File টি R- এ Save থাকবে এবং যখন ইচ্ছে এখানে edit করতে পারবেন। এখন কোনো ফাইল কম্পিউটার থেকে হারিয়ে যাবার ভয় Introduction This is the first R Book in Bangle.

৭৪ ● সহজ বাংলায় R প্রোগ্রামিং

The Objective is to help effective R learning. Goal Goal is to help R learners in Bangladesh.

• ------

আপনি দুইভাবেই Practice করতে পারেন। এখানে Output চিত্র -১

এর মত আসবে। এবার, আমরা নিজেরা আরেকটি উদাহরণ data

(wagel, package = 'wooldridge') দিয়ে Practice করি

এখানে Wage1 = dependent variable এবং Education =

মেশিন লার্নিং অ্যালগরিদম ব্যবহার করে আমরা খুব সহজেই Data

Mining, Analysis, Prediction এবং বর্তমান যুগের অনেক

Business related সমস্যার সমাধান করতে পারি। নিচের Step

গুলোতে মেশিন লার্নিং অ্যালগরিদম ব্যবহার করার আগে আমাদের কি

Add In-text Citations / References to Your Report

মেশিন লার্নিং অ্যালগরিদম এবং R এর ব্যবহার

করতে হবে বা জানতে হবে তা দেখানো হয়েছে :

independent vartiable |

৫৮ 🌘 সহজ বাংলায় R প্রোগ্রামিং

0 10 20 30 40 50

R- Shiny Application Package আপনি যদি Microsoft Power Bi এর মতো একটা Data Visualization Dashboard করতে চান অথবা Interactive Web Application develop করতে চান সেক্ষেত্রে R- Shiny Application Package হচ্ছে একটি Perfect Platform। Re Studio তে R- Shiny package টি Install করে আপনি খুব সহজেই একটি Interactive Web Application তৈরি করতে পারবেন। R- Studio Shiny Lesson এ এই বিষয় সম্পর্কে বিস্তারিত রয়েছে। আপনি এই Lesson টি step by step follow করে আপনার প্রথম Web Application তৈরি করতে পারবেন। নিচের চিত্র-২২ এবং ২৩ এর মতো করে।

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৭৫

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৫৯

Deploy Model

Is the model run manually on

Monitor Performance shecdule? Is it deployed stand-alone or integrated into current

সাধারণত মেশিন লার্নিং অ্যালগরিদম দু'ভাগে বিভক্ত। সুপারভাইজঙ

লার্নিং এবং আনসুপারভাইজড লার্নিং। সুপারভাইজড লার্নিং মূলত

Prediction বা অনুমান নির্ভর মডেল নিয়ে কাজ করে। অন্যদিকে

আনসুপারভাইজড লার্নিং আমাদের ডাটা প্যাটার্নের ব্যাপারে বুঝতে

সহায়তা করে। মেশিন লার্নিং ডাটা সায়েন্সের modeling অংশের

প্রয়োগ এবং এটি মূলত Big Data নিয়ে কাজ করার সময় বেশি

ব্যবহৃত হয়। Big data সাধারণত এমন ডাটা সেট যা কয়েক শ থেকে

কয়েক মিলিয়ন হতে পারে। উপরের চিত্রটি লক্ষ্য করলে আমরা বুঝতে

পারি যে মেশিন লার্নিং বা ইকোনোমেট্রিক্সের মডেল কোনো প্রজেক্টে

ব্যবহারের আগে ডাটার ধরন বুঝতে পারাটা খুবই জরুরি। এবার আসা

যাক, সুপারভাইজড এবং আনসুপারভাইজড লার্নিং মডেলের ব্যবহার

নিয়ে। সুপারভাইজড লার্নিং মানে হচ্ছে একজন Programmer

মডেলকে গাইড করে ডাটার ব্যবহার শেখাতে সক্ষম। কারণ ডাটার

চিত্ৰ-২২

Hello Shiny!

Number of bins:

1 30

1 6 11 16 21 26 31 36 41 46 50

ডাটার ব্যবহারের জন্য কিছু গুরুত্বপূর্ণ Source এই বইয়ের বেশির ভাগ উদাহরণ আপনি ফ্রী ডাটা Source দিয়ে Practice করতে পারেন। এই Source গুলো থেকে ডাটা ফ্রী এবং সহজেই ডাউনলোড করা যায়। Kaggle, Our World in Data, WDI ইত্যাদি ডাটা Source গুলো সম্পূর্ণ ফ্রী এবং আপনি যেকোনো সময় ডাটা ডাউনলোড করে এই বইয়ের প্রত্যেক অধ্যায়ের কোড বা কমাভণ্ডলো Practice করতে পারবেন দেশ ভিত্তিক, কোম্পানি ভিত্তিক ডাটা দিয়ে। পরবর্তী সংস্করণে আমি R- প্রজেক্ট ভিত্তিক বিষয় নিয়ে লিখবো এবং ডাটা সায়েন্সের সব প্রজেক্টের Concept নিয়ে বিস্তারিত আলোচনা করবো। আমার এই বইয়ের বিষয়গুলো R- প্রোগ্রামিংয়ের Introductory level cover করবে। Advance level নিয়ে পরবর্তী বইয়ে আলোচনা করবো।

৭৬ 🌘 সহজ বাংলায় R প্রোগ্রামিং

install.packages("rattle")

library(rattle)

rattle( )

ব্যাস এইটুকু!! বাকি সব কাজ বা অ্যানালাইসিস আপনি Rattle এ

করতে পারবেন খুবই সহজে। আগ্রহী পাঠকগণ Data Mining with

Rattle and R by Graham Williams এর বইটি পড়তে পারেন।

R- ছাড়া, Python, Weka, Orange, Rapid Miner Software এ

te New Open Save Report Export Stop Quit

Input Input

er 🖰 audit.csv 🔝 Separator: 🗋 Decimal: 🗍 🗹 Header

Partition 70/15/15 Seed: 42 💆 View Edit.

৬০ 🌘 সহজ বাংলায় R প্রোগ্রামিং

মেশিন লার্নিং মডেল নিয়ে কাজ করা যায়।

প রি শে ষ

এটা আমার প্রথম বই। এই বইটি তাদের জন্য যারা নতুন করে R বা Python শিখার জগতে প্রবেশ করতে চাচ্ছে এবং বিভিন্ন ব্যাকগ্রাউন্ড থেকে এসেছেন। বইটিতে আমি কিভাবে শিখেছি, কতটুকু শিখে কাজ শুরু করেছি গত এক বছরে তাই আপনাদের সাথে শেয়ার করলাম। এই নতুন পৃথিবী দক্ষ মানুষ চায়, আর নিজেকে R বা Python এর মতো কম্পিউটারের ভাষা শিখানো আমার কাছে খুবই রোমাঞ্চকর অভিজ্ঞতার মনে হয়েছে। আগামী দিনে কম্পিউটার এর ভাষা বুঝা, তাকে গাইড করা, বড় বড় ডাটা নিয়ে কাজ করা এবং ব্যবসা সংক্রান্ত সমস্যাগুলো ডাটা নির্ভর অ্যালগরিদম দিয়ে সমাধান করা মানুষের অনেক বেশি প্রয়োজন হবে। আর আমি একা এই নতুন পৃথিবীতে প্রবেশ করতে চাই না, একা নতুন এই দুনিয়ার ভাষা শিখতে চাই না, আপনাদেরও নিয়ে যেতে চাই। আমরা যদি সবাই একই ভাষায় কথা বলি, তবেই না আড্ডা জমবে, তাই না??

Big Data + Machine learning = Data Science

আর এর সাথে অর্থনীতি, ইকোনোমেট্রিক্স এবং ব্যবসা সংক্রান্ত চমৎকার

সমস্যা ও সমাধান যোগ করলে হবে Business Data Science

Big Data মানে ডাটার আওতা বা Size এতো বিশাল যে কোন একটি

কম্পিউটার অনেক সময় ধারণ করা কঠিন হয়ে পড়ে। যেমন,

Facebook Data, E-mail Data অথবা আমাদের দেশের NID

তথ্য ইত্যাদি। এই Complex ডাটা থেকে Machine learning

algorithms (অ্যালগরিদম) Predicition Performance খুব সহজে

বের করতে পারে। আপাতত এতটুকু জানলে হবে, পরবর্তীতে এ নিয়ে

বিস্তারিত লেখার ইচ্ছা আছে. যাতে আমি প্রকত ব্যবসা সমস্যা কিভাবে

সমাধান করতে হয় Machine learning algorithms দিয়ে ত

আলোচনা করব। এবার Python এবং R- এর Syntax বা গঠনগত

দিকের কিছু সাদৃশ্য বৈসাদৃশ্যের উদাহরণ দিব। আগেই বলেছি R-

একটি ফাংশন (Function-based) ল্যাংগুয়েজ, Python সেখানে

একটি General Purpose ল্যাংগুয়েজ। Python এ ফাংশন রয়েছে

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৬১

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৭৭

তবে R এর মত এতটা নির্ভর নয়।

'type' বলে,

a < - c (4,5,1,3,2) R এর ক্ষেত্রে print (class (a))

Python G, a = [4, 5, 1, 3, 2] print (type (a))

R এর কোন Function এ help জানতে চাইলে (?plot) এই কমাভটি

দিতে হয় এবং Python এ জানতে চাইলে help(plot) এই কমাভটি

দিতে হয়।

R- এর অনেক Data Set বা Data Frame (df) R- install করার

সাথে Built in থাকে। যদি তা দেখতে চান, তাহলে data( ) কমান্ড

দিলেই দেখতে পাবেন Built in Data R-studio তে। Python

Scikit Package এ এই ধরনের Built in free Data থাকে। যা

Jupiter Console (R- studio এর মত) উদাহরণ সরূপ import

sklearn as sklearn, import pandas as pd এবং data =

datasets. load-iris( ) এর মতো করে দেখা যেতে পারে। iris

নামের ডাটা সেটি free load করা যায়। এ ধরনের আরো কিছু সাদৃশ্য

এবং বৈসাদৃশ্য রয়েছে যা আমরা দ্রুতই জানবো কাজ করতে করতে

৬২ 🌘 সহজ বাংলায় R প্রোগ্রামিং

Facebook Page

৭৮ 🌘 সহজ বাংলায় R প্রোগ্রামিং

Skill Tone

উপরের চিত্র-০৬ একটু লক্ষ করি, এখানে Time Variable (Quarters) এমনভাবে বিভিন্ন কলামে বিন্যস্ত বা Distributed হয়েছে যে, প্রত্যেকটি যেন এক একটি Independent Variable হিসেবে Appear হয়েছে। আমরা কিন্তু tidyr package টি দিয়ে সুন্দরভাবে Year wise Quarter গুলোকে সাজাতে পারি ম্যাজিকের মতো। উপরের চিত্র-০৭ টি একটু খেয়াল করলেই ব্যাপারটা পরিষ্কার

ধরনের Statistical Analysis R-এ শুরু করা যাবে। এই পদ্ধতিতে ডাটা অ্যানালাইসিস করাকে GUI (Graphical User Interface Based Work) বলে। R-Commander ছাড়াও Rattle Package আরেকটি জনপ্রিয় GUI যা দিয়ে খুব সহজেই মেশিন লার্নিং মডেল Analysis করা যায়।

প্রাথমিকভাবে R-Commander install করার Command হচ্ছে install.packages ("Rcmdr", dependencies = TRUE) library(Rcmdr) এখানে dependencies = TRUE মানে হচ্ছে Remdr R এর অসংখ্য প্যাকেজ ব্যবহার করে আপনার Command কে execute করে। এর মানে R Commander Package অন্যান্য R Package গুলোর উপর নির্ভরশীল। তাই আমরা লিখলাম dependencies = যদি Remdr বা Rattle দিয়ে খুব সহজে সব Command execute বা run করা যায়, তাহলে R-এর console এ এতো এতো Command লিখার দরকার কি? আসলে আমরা যদি R-এর Basic

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৬

A difficial set.

A difficial R Variable editor R Variable editor ৬৪ 🌘 সহজ বাংলায় R প্রোগ্রামিং

চিত্ৰ-১১: Entering data using R Commande

সহায়ক গ্রন্থ ও ওয়েবসাইট 1. cran.r-project.org 2. Business Data Science by MATT TADDY 3. Practical Machine learning with R and Python by T.V 4. Using R for IntroductoryEconometrics 6. Kaggle Data, Our World in Data, WDI Wooldridge Econometrics Book Data. আমার Youtube Channel Dr. Munshi Naser- Skill Tone Facebook Group Skill Learns Group with Dr. Munshi

সহজ বাংলায় R প্রোগ্রামিং 🌘 ৭৯

৮০ 🌒 সহজ বাংলায় R প্রোগ্রামিং