

چرا و چگونه زبان R به توسعه کسب و کار کمک می کند

بسیاری از کارشناسان بر این باورند که علم داده برای کسب و کارها (DS4B) آینده تحلیل تجاری است اما اغلب افراد از ابزارها و

روش های لازم در این زمینه اطلاع ندارند. مهم ترین نکته ای که باید از آن اجتناب کنید هدر دادن زمان با استفاده از ابزارهای

ناکارآمد است. استفاده موثر از زمان شامل دو بخش است: (۱) انتخاب ابزارها متناسب با کسب و کار (۲) فراگیری کامل روش کار با

این ابزارها برای دستیابی به ارزش

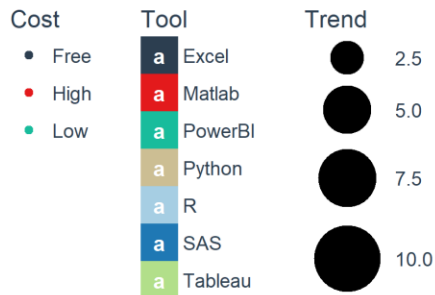
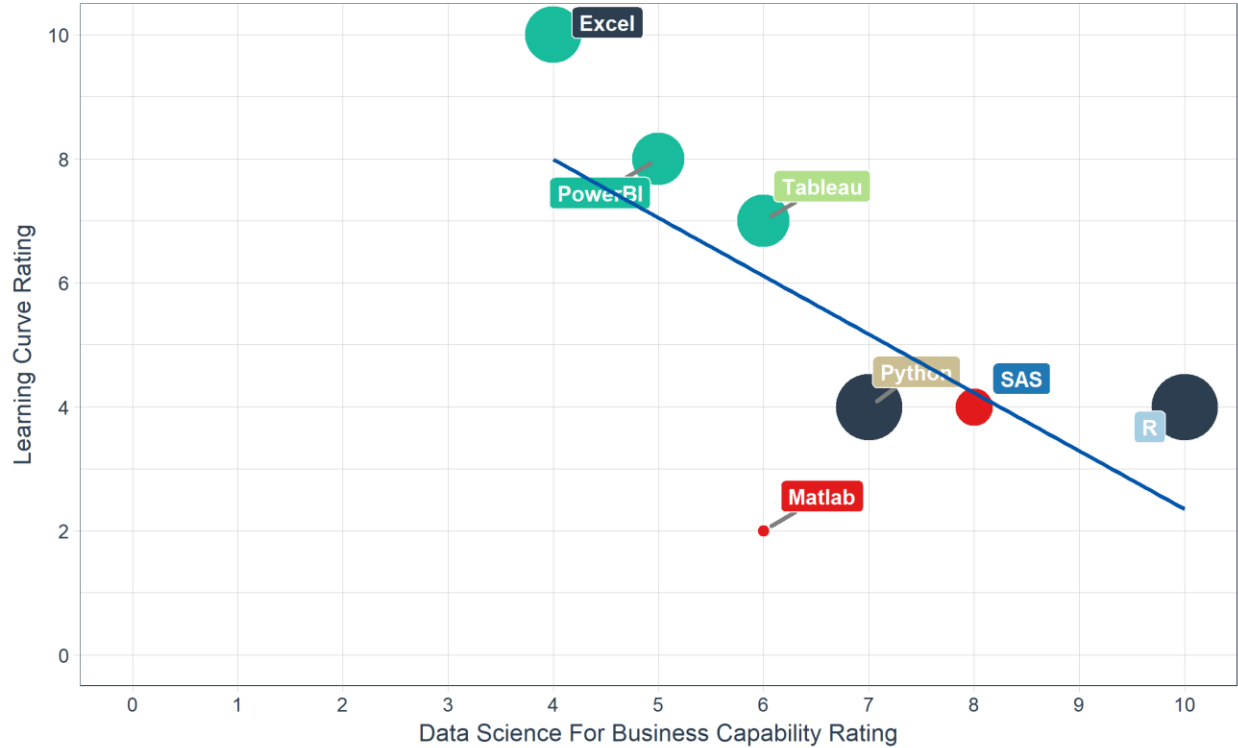
این مقاله با تمرکز روی بخش اول، ۶ دلیل را برای اثبات کارآمد بودن زبان برنامه نویسی R به عنوان ابزاری مناسب برای تحلیل،

ارائه کرده است.

دلیل اول R از بالاترین سطح کلی کیفیت برخوردار است

DS4B Tools: Capability Vs Learning Curve

R has a longer learning curve but has a massive business capability rating



Why R? Tools like Excel, Tableau, PowerBI are easier to learn, but have lower Business Capability. Tools like Python, SAS, and Matlab have high Data Science Capability, but lack the visualization and interactive application tools needed for business. R has the best data science, visualization, and interactive tools plus it's free!

DS4B – Data Science for Business ابزارهای مختلفی برای تحلیل یا هوش کسب و کار در دسترس قرار دارند و

نیز زیر مجموعه ای از آنها به شمار می رود. هر کدام از این ابزارها مزایا و معایب خاص خود را به همراه دارند که ممکن است روی

کسب و کار تاثیر مهمی داشته باشند. در ادامه برخی از ابزارهای شناخته شده تر را از جوانب مختلف با هم مقایسه کرده ایم:

- قابلیت کسب و کار (کم=۱، زیاد=۱۰)
- سهولت یادگیری (دشوار=۱، آسان=۱۰)
- هزینه (رایگان/حداقل، کم، زیاد)
- استقبال کاربران (کاهش سریع=۰، ثابت=۵، رشد سریع=۱۰)

نتیجه به دست آمده بسیار جالب بود. بسط محاسبه شده از سری داده ها نشان دهنده موازنه ای بین منحنی یادگیری و نرخ قابلیت های DS4B است. در واقع یادگیری انعطاف پذیرترین ابزارها که از بیشترین قابلیت های کسب و کار برخوردار هستند، از بقیه دشوارتر است. در عین حال ابزارهای با یادگیری آسان معمولاً قابلیت های علوم داده و کسب و کار چندانی را به ارمغان نمی آورند. از نظر ما بهتر است قابلیت های ابزار را نسبت به سهولت یادگیری در اولویت قرار دهید.

در میان ابزارهای بررسی شده زبان R به لطف برخورداری از قابلیت های گسترده و متنوع کسب و کار علوم داده از یک سو و قیمت پایین از سوی دیگر به سرعت در حال رشد است. تنها نقص این زبان را می توان منحنی یادگیری آن به شمار آورد. در ادامه این مقاله به بررسی مزایای R برای کسب و کارها پرداخته ایم.

دلیل دوم R یک ابزار علوم داده برای متخصصان خارج از حوزه کامپیوتر است

اگر به دنبال یک ابزار علوم داده با کارایی بالا هستید، دو گزینه پایتون یا R را به شما پیشنهاد می کنیم. کار کردن با هر دو گزینه در شروع کار اشتباه است و بهتر است با توجه به نیازهایتان یکی از آنها را انتخاب کنید. تفاوت های بین پایتون و R در دیتابیس ها و اینفوگرافی های بسیاری شرح داده شده اما آنچه که اغلب مورد غفلت واقع شده تناسب با حوزه فعالیت شخص است که در ادامه دلایل آن را شرح داده ایم.

۱. اکثر افراد علاقمند به یادگیری علوم داده به خاطر کسب و کار در حرفه‌هایی نظیر مکانیک، شیمی و غیره حرفه‌ای هستند اما در زمینه کامپیوتر یا نرم افزارهای تحلیل تخصصی ندارند.
 ۲. بخش عمده‌ای از فعالیت‌های تجاری و مالی شامل ارتباطات است. این فعالیت‌ها در قالب گزارش، مجموعه اطلاعات و اپلیکیشن‌های تعاملی تحت وب صورت گرفته و به تصمیم‌گیرندگان اجازه می‌دهد وقوع هرگونه مشکل یا چالش را شناسایی کرده و با اتخاذ تصمیمات آگاهانه کسب و کار را به مسیر درست هدایت کنند.
- پس از آشنایی با این موارد مهم باید به سراغ دو بازیگر اصلی در حوزه علوم داده برویم.

پایتون

پایتون یک زبان برنامه‌نویسی چند منظوره است که توسط مهندسی نرم افزار توسعه یافته و از کتابخانه برنامه‌نویسی غنی برای ریاضی، آمار و یادگیری ماشین برخوردار است. این زبان از بهترین ابزارها برای یادگیری ماشین و یادگیری عمیق برخوردار است اما از فقدان زیرساخت‌های لازم برای موضوعاتی از قبیل گزارش‌گیری برای ارتباطات یا مسائل اقتصادی رنج می‌برد. به همین خاطر پایتون بیشتر مناسب مهندسان نرم افزار و متخصصان علوم کامپیوتر است.

زبان برنامه‌نویسی R

زبان برنامه‌نویسی و محیط نرم‌افزاری آماری R نیز از کتابخانه‌ای غنی برای علوم داده، یادگیری ماشین و آمار بهره می‌برد. R به لطف پکیج‌های اختصاصی برای موضوعات مختلف و زیرساخت ارتباطی، برای کسب و کارها بسیار مناسب است. پکیج‌های R دامنه‌ای گسترده‌ای از موضوعات مختلف نظیر اقتصاد، امور مالی و سری‌های زمانی را پوشش می‌دهند. علاوه بر این ابزارهای قدرتمند برای تصویرسازی، گزارش‌گیری و تعامل که برای کسب و کارها اهمیتی برابر با علوم دارند، از دیگر ویژگی‌های شاخص این زبان هستند. قابلیت‌های مذکور زبان برنامه‌نویسی R را به انتخابی مطلوب برای دانشمندان، مهندسان و حرفه‌ای‌های کسب و کار تبدیل کرده است.

کدام یک برای شما مناسب‌تر است؟

برای انتخاب بین این دو زبان باید به سوالات زیر پاسخ دهید:

- آیا شما یک متخصص کامپیوتر یا مهندس نرم افزار هستید؟ اگر بله پس سراغ پایتون بروید.
- آیا شما یک تحلیلگر حرفه ای یا مهندس فعال در حوزه شیمی، مکانیک یا صنایع دیگر هستید؟ اگر بله پس R را یاد بگیرید.

هدف شما از فراگیری این زبان ها چیست؟

- آیا به دنبال تولید یک ماشین خودران هستید؟ اگر بله پس پایتون را انتخاب کنید.
- آیا هدفتان انتشار تحلیل های کسب و کار در سراسر سازمان است؟ اگر بله سراغ R بروید.

دلیل سوم: یادگیری R با پکیج TidyVerse آسان است

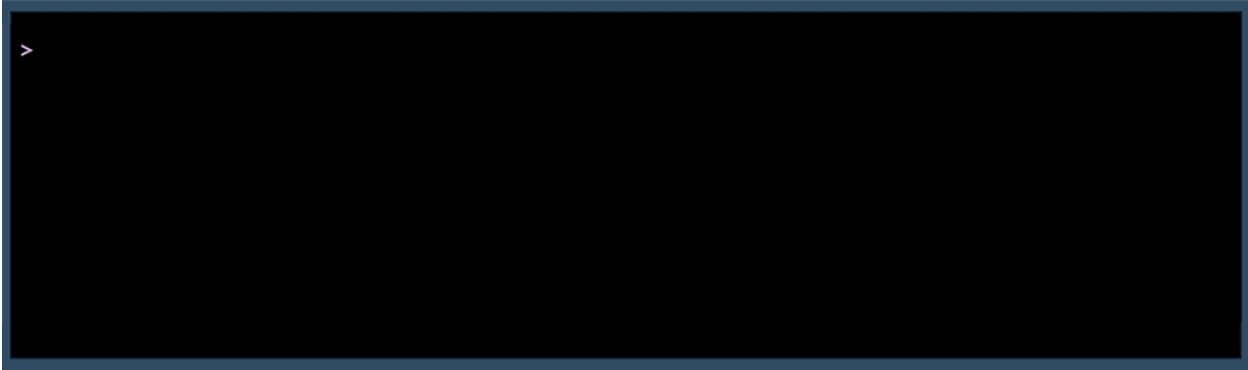
پیش از این تسلط بر زبان R بسیار دشوار بود چرا که برخلاف زبان های دیگر در آن ساختار، اولویت نداشتند و به همین خاطر R پیچیده و متناقض بود. معرفی مجموعه ابزارها و پکیج های TidyVerse با واسط کاربری برنامه نویسی ساختاریافته، این روند را تغییر داد.

ابزارهایی مثل `dplyr` و `ggplot2` با ارائه راهکارهایی ساختارمند و پایدار برای کار با داده ها یادگیری این زبان را آسان تر از قبل

کردند. در ادامه تلاش های افرادی مثل هدلی ویکهام (Hadley Wickham) پکیج های TidyVerse با دربرداشتن

مجموعه از ابزارهای کار با داده، تصویرسازی، بازگویی، مدل سازی و ارتباطات به یادگیری آسانتر زبان برنامه نویسی R کمک

کردند.



کارشناسان با تولید بسته‌های جدید مبتنی بر زیرساخت **TidyVerse** به تکامل هرچه بیشتر زبان برنامه‌نویسی و محیط نرم افزاری **R** کمک می‌کنند. تمرکز این افراد بیشتر روی مدل‌سازی و الگوریتم‌ها است؛ علاوه بر این **TidyVerse** با توسعه هرچه بیشتر حالا مباحثی از قبیل متون (**tidytext**) و امور مالی (**tidyquant**) را نیز پوشش می‌دهد. تلاش‌هایی از این دست نشان‌دهنده آینده روشن **R** است و این مساله جسارت و اعتماد به نفس بیشتری را به افراد علاقمند به یادگیری این زبان می‌دهد.

دلیل چهارم **R** دارای قلب، مغز و عضله است

اکتفا کردن به اینکه **R** یک زبان قدرتمند است نمی‌تواند تمامی جوانب آن را به رخ بکشد. در واقع باید گفت که این زبان از نظر کسب و کار چیزی در حد اکسل تقویت شده با هوش بیشتر است.

مغز **R**

این زبان الگوریتم‌های پیشرفته‌ای از قبیل موارد زیر را پیاده‌سازی می‌کند:

- **H2O** : پکیج پیشرفته یادگیری ماشین

- Keras/TensorFlow (کراس، تنسورفلو) : بسته های یادگیری عمیق تخصصی

- xgboost : بهترین الگوریتم Kaggle

- و موارد بسیاری دیگری

این ابزارها در زمینه های مختلف از محصولات هوش مصنوعی گرفته تا مسابقات Kaggle کاربرد داشته و می توان از آنها در

تحلیل کسب و کار بهره برد

عضلات R

این زبان از ابزارهای قدرتمندی برای موارد زیر برخوردار است:

- عملیات بُرداری: استفاده از عملیات بُرداری برای انجام سریع محاسبات ریاضی

- حلقه ها (purrr)

- عملیات موازی (parallel, future)

- افزایش سرعت اجرای کدها با استفاده از C پلاس پلاس (Rcpp)

- اتصال به زبان های دیگر (rJava, reticulate)

- کار با دیتابیس ها و اتصال به آنها (dbplyr, odbc, bigquery)

- مدیریت داده های کلان و اتصال به آپاچی اسپارک (sparklyr)

- و موارد بسیار دیگر

قلب R

پیشتر درباره زیرساخت ها و **tidyverse** که امکان تولید اکوسیستم اپلیکیشن ها با استفاده از رویکردی ثابت را می داد صحبت کردیم. تجزیه و تحلیل کارآمد داده ها نیز مدیون همین زیرساخت است که موارد زیر را میسر می سازد:

- دستکاری داده ها (**dplyr, tidyr**)
- کار با انواع مختلف داده ها: **stringr** برای رشته ها، **lubridate** برای زمان و تاریخ، **forcats** برای فاکتورها و

موارد مطلق

- تصویرسازی (**ggplot2**)
- برنامه نویسی (**purrr, tidyeval**)
- ارتباطات (**Rmarkdown, shiny**)

دلیل پنجم R برای کسب و کار ساخته شده است

دو مزیت اصلی **R** نسبت به دیگر زبان های برنامه نویسی این است که با استفاده از آن می توان اپلیکیشن های مبتنی بر یادگیری ماشینی تحت وب و گزارش های آماده کسب و کار را تولید کرد. نه پایتون و نه **Tableau** در این زمینه ها کارایی مشابه با **R** ندارند. این دو قابلیت عبارتند از **RMarkdown** برای تولید گزارش و **shiny** برای اپلیکیشن های تعاملی وب.

چارچوب **RMarkdown**

Rmarkdown یک چارچوب ایجاد گزارش های قابل تکثیر است که امروزه در زمینه تولید بلاگ، ارائه، وب سایت، کتاب، ژورنال و غیره نیز کاربرد دارد. با استفاده از این ابزار می توان کدها را همراه با متن قرار داده و امکان مشاهده تحلیل، خروجی و توضیحات را برای همه افراد فراهم کرد. این فناوری تاکنون تکامل بسیاری پیدا کرده و مثال هایی از قابلیت های آن عبارتند از:

- **rmarkdown** برای تولید گزارش های **HTML**، **ورد** و **PDF**
- **rmarkdown** برای ارائه
- **flexdashboard** برای ایجاد اپ های وب از طریق فرمت کاربر پسند **Rmarkdown**
- **blogdown** برای ایجاد بلاگ و وبسایت
- **bookdown** برای ایجاد کتاب های آنلاین
- اسناد تعاملی
- تولید گزارش های پارامتربندی شده و سفارشی (برای مثال گزارش هایی برای یک بخش جغرافیایی، دپارتمان یا محدوده زمانی مشخص)

چارچوب SHINY

SHINY یک چارچوب برای ایجاد اپ های تعاملی تحت وب از طریق **R** است. این ابزار نه تنها قدرتمند است بلکه به واسطه ابزارهای تصمیم گیری تعاملی به متخصصان خارج از علم داده نیز امکان بهره گیری از داده ها را می دهد. در تصویر زیر یک اپ گوگل ترند را می بینید که با استفاده از **SHINY** طراحی شده است.

دلیل ششم: انجمن های پشتیبانی بی نظیر **R**

قدرتمند بودن یک زبان به خودی خود کافی نیست و برای موفقیت به جامعه پشتیبان نیاز دارد. خوشبختانه **R** به واسطه **CRAN** و **R Community** در این زمینه هم عملکردی مطلوب دارد.

CRAN: بسته های **R** تهیه شده توسط عموم

CRAN را می توان فروشگاه آنلاینی مثل اپ استور در نظر گرفت که محتوای ارائه شده در آن رایگان، بسیار کارآمد و ویژه R

هستند. این مارکت با بیش از ۱۴ هزار پکیج ارائه شده تمام نیازهای شما در زمینه های یادگیری ماشین، محاسبات کارآمد، امور

مالی و اقتصادی را برطرف خواهد کرد. یکی از راه های بررسی پکیج ها بخش **task views** است که زمینه های خاص را

پوشش می دهد.

CRAN توسط عموم کاربران پشتیبانی شده و نویسندگان شناخته شده در حوزه محتوای منبع باز از جمله هدلی ویکهام و

دریک ادلبوتل آن را هدایت می کنند. توسعه پکیج روش بسیار خوبی برای مشارکت در این جامعه و نمایش مهارت کدنویسی

است.

انجمن های پشتیبانی

اگر دلیل روی آوردن شما به R قابلیت های آن باشد، دلیل ادامه دادن این مسیر نیز احتمالاً انجمن های آن خواهد بود

R Community یکی از بهترین بخش های این زبان برنامه نویسی است که در آن افراد مستعد و متخصص در محیطی

صمیمی و دوستانه گرد هم جمع شده اند.

شبکه های اجتماعی و وب

کاربران R در همه جای وب حضور دارند و برخی از محبوب ترین محل های تجمع آنها عبارتند از:

- R-Bloggers
- هشتگ #rstats در توییتر
- گروه پروژه R برای محاسبات آماری در لینکدین

کنفرانس ها

- کنفرانس های کسب و کار متمرکز بر R توجهات زیادی را به خود جلب می کنند. در ادامه به برخی از معروف ترین رویدادها در این زمینه اشاره شده است:

- EARL : کنفرانس مبتنی بر راه حل های مانگو برای اپلیکیشن های کسب و کار و سازمانی R

- R/Finance : کنفرانسی درباره دارایی مالی، تحلیل نمونه کارها و امور مالی کاربردی

- Rstudio Conf : کنفرانس فناوری Rstudio

- New York R : کنفرانس متمرکز بر فناوری و کسب و کار R

با کلیک روی این لینک می توانید به لیست کاملی از کنفرانس های R دسترسی پیدا کنید.

جلسات و ملاقات های غیر رسمی

یکی از نکات جالب درباره R این است که در بسیاری از شهرها ملاقات هایی درباره این فناوری صورت گرفته و در آن کاربران این

زبان دور هم جمع شده و درباره جوانب مختلف R به گفتگو می پردازند. این قرار ملاقات ها معمولاً توسط R-Consortium

سازماندهی می شوند. فهرست کاملی از آنها را می توانید در این لیست مشاهده نمایید.

نتیجه گیری

R به لطف برخورداری از مزایای گسترده به انتخابی مطلوب برای علوم داده در کسب و کارها (DS4B) بدل شده است. واضح

است که پایتون نیز از قدرت بالایی برخوردار است اما با توجه به نیازهای کسب و کارها که در این مقاله توضیح داده شد، توان

رقابت با R را ندارد.