



چگونه تحلیل داده با پایتون را یاد بگیریم؟



تولید شده توسط تیم محتوا سماتک

www.sematetec-co.com

چگونه تحلیل داده با پایتون را یاد بگیریم؟



با ورود به عصر داده، نیاز به تجزیه و تحلیل داده ها نیز افزایش یافت. همچنین دست یابی به موفقیت و سود بیشتر از طریق تحلیل داده به یکی از اصلی ترین چالش ها برای صنایع مختلف تبدیل شد. در این زمان، علم داده (Data Science) به صنایع و سازمان ها کمک کرد تا بتوانند درآمد بیشتری کسب کنند. اما پیدایش و توسعه علم داده نیازمند ابزارهایی برای پیاده سازی و استفاده از مفاهیم آن در جهان واقعی است. از همین رو نرم افزارها و زبان های برنامه نویسی بسیاری مانند پایتون توانستند جایگاه خوبی را در میان پژوهشگران و علاقمندان به این حوزه کسب کنند. بنابراین، در ادامه به شما خواهیم گفت که چگونه تحلیل داده با پایتون را یاد بگیرید.

آشنایی با علم داده

علم داده (Data Science) ترکیبی از ابزارهای مختلف، الگوریتم ها و اصول یادگیری ماشین با هدف کشف الگوهای پنهان از داده های خام است. برای درک بهتر این موضوع، فیلم های علمی تخیلی هالیوود را در نظر بگیرید. تمام ایده های این گروه از فیلم ها می تواند توسط علم داده به واقعیت تبدیل شود. بنابراین، با کمک تحلیل داده و هوش مصنوعی، افراد می توانند تصمیم های بهتری در حوزه ی کسب و کار خود بگیرند.

پایتون چیست؟

پایتون (Python) یک زبان برنامه نویسی محبوب می باشد که توسط خیدو فان روسوم طراحی و در سال 1991 منتشر شده است.

از این زبان برنامه نویسی سطح بالا برای توسعه وب (سرور)، توسعه نرم افزار، ریاضیات و برنامه نویسی سیستم استفاده می شود.

با توجه به این نکته که پایتون یک محیط پویا برای برنامه نویسی دارد و همچنین کاملاً رایگان و اپن سورس است، از آن برای تحلیل داده نیز استفاده می شود.

چرا باید برای تحلیل داده پایتون را یاد بگیریم؟



درک پایتون یکی از مهارت های مهم و ضروری، برای بدست آوردن موفقیت در بازار کار تحلیل داده است. اما قبل از این که نحوه یادگیری پایتون برای تحلیل داده را بررسی کنیم، به طور خلاصه توضیح می دهیم که چرا تحلیل داده با پایتون بسیار مورد توجه قرار می گیرد. زیرا:

- در سال 2016 ، در کگل (پلت فرم برتر مسابقات علوم داده)، زبان برنامه نویسی پایتون از زبان برنامه نویسی R سبقت گرفت.
- در سال 2017، در نظرسنجی ابزارهای پرکاربرد دانشمندان داده که سالانه توسط سایت KD Nuggets انجام می شود، زبان پایتون توانست زبان R را پشت سر بگذارد.

• در سال 2018 ، 66٪ از دانشمندان تحلیل داده، استفاده روزانه از پایتون را گزارش کردند و این موضوع باعث شد که زبان پایتون، به عنوان زبان اول متخصصان تجزیه و تحلیل داده انتخاب شود.

کارشناسان علوم داده انتظار دارند این روند با رشد روزافزون در اکوسیستم پایتون ادامه یابد. اگرچه ممکن است سفر شما برای یادگیری برنامه نویسی پایتون تازه آغاز شده باشد، اما خوب است بدانید که فرصت های شغلی فراوان و در حال رشد در این حوزه برای شما وجود دارد.

همچنین میانگین حقوق و دستمزد یک دانشمند داده 121،583 دلار در سال است و انتظار می رود که این مقدار افزایش یابد، زیرا نیاز به دانشمندان داده در صنایع مختلف در حال رشد می باشد و در آینده، تقاضا برای دانشمندان داده به شدت از عرضه پیشی می گیرد.

بنابراین، آینده برای علم داده روشن است و یادگیری تحلیل داده با پایتون می تواند به شما در دست یابی به موفقیت در این زمینه کمک کند. خوشبختانه، یادگیری پایتون و سایر اصول برنامه نویسی مانند همیشه قابل دستیابی است. ما در پنج مرحله به شما خواهیم گفت که چگونه از پایتون برای تحلیل داده استفاده کنید.

اما به یاد داشته باشید، ساده بودن مراحل به این معنی نیست که مجبور نیستید، کاری انجام دهید. از طرفی، اگر وقت کافی را به پایتون اختصاص دهید، نه تنها مهارت جدیدی را یاد می گیرد، بلکه به طور بالقوه در زمینه ی فعالیت کاری خود به سطح بالاتری دست خواهید یافت.

مرحله 1: یادگیری اصول پایتون (Python) برای تحلیل داده

اولین قدم برای شما، یادگیری اصول برنامه نویسی پایتون خواهد بود. همچنین باید به مرور با علم داده نیز آشنا شوید.

یکی از بهترین ابزارهایی که در آغاز سفر تحلیل داده با پایتون به شما کمک زیادی می کند، مطالعه ی کتاب های HeadFirst می باشد. می توانید فایل pdf این کتاب ها را در انتهای مقاله دانلود کنید.

برای شروع آشنایی با علم داده، می توانید عضو گروه ها و شبکه های اجتماعی شوید که در این زمینه فعالیت می کنند.

با پیوستن به این گروه ها، در کنار افراد همفکر خود قرار می گیرید و فرصت های شغلی خود را افزایش می دهید. طبق تحقیقات انجام شده، 30٪ از کل استخدام های نیروی جدید، با معرفی کارمندان انجام می شود.

برای آشنایی بهتر با علم داده در سایت کگل (Kaggle) یک حساب کاربری ایجاد کنید و مطالب آن را مطالعه کنید.

یک مهارت ضروری برای تحلیل داده با پایتون

یکی از مهارت هایی که می تواند به شما در حوزه علم داده کمک زیادی کند، شناخت رابط خط فرمان (CLI) می باشد. رابط خط فرمان (CLI) به شما این امکان را می دهد که اسکریپت ها را با سرعت بیشتری اجرا کنید و سریعتر برنامه ها را آزمایش کنید، در نتیجه می توانید با داده های بیشتری کار کنید.

مرحله 2: انجام پروژه های کوچک پایتون

ما واقعاً به یادگیری عملی ایمان داریم. شاید تعجب کنید که به زودی برای ساخت پروژه های کوچک پایتون آماده خواهید شد.

شما می توانید برنامه های کوچک مانند: ماشین حساب، بازی آنلاین ساده و انواع پروژه های ساده پایتون را در گوگل جست و جو و برنامه نویسی کنید تا با پایتون بیشتر آشنا شوید.

پروژه های کوچک برنامه نویسی برای همه زبان ها استاندارد است و راهی عالی برای تقویت درک شما از اصول اولیه هستند.

همچنین در این مرحله از آموزش تحلیل داده با پایتون باید در مورد API ها و web scraping شناخت پیدا کنید. web scraping علاوه بر این که به شما در یادگیری برنامه نویسی پایتون کمک می کند، در آینده نیز برای جمع آوری اطلاعات مفید خواهد بود.

خواندن منابع جدید در حوزه تحلیل داده با پایتون

تکالیف دوره خود را افزایش دهید و پاسخ چالش های برنامه نویسی پایتون را پیدا کنید. کتاب های راهنما، پست های وبلاگ و حتی کدهای اوپن سورس دیگران را بخوانید تا بهترین روش ها برای تحلیل داده با پایتون را بیاموزید و ایده های جدید دریافت کنید. همچنین در ادامه 3 کانال برتر YouTube برای تحلیل داده بر اساس تعداد مشترکان را به شما معرفی می کنیم.

1. ادورکا (edureka)

مشترکین: 2,440 K ، فیلم ها: 4012

در این کانال مباحث علم داده به شکل ساده و روان تدریس می شود. در واقع ادورکا یک پلتفرم یادگیری الکترونیکی زنده و تعاملی است که به شما کمک می کند به صورت رایگان با تحلیل داده آشنا شوید.

2. جوما تک (Joma Tech)

مشترکین: K 889 ، فیلم ها: 83

در این کانال در مورد زندگی در سیلیکون ولی، شرکت های بزرگ فناوری، علوم داده و مهندسی نرم افزار صحبت می شود.

3. یادگیری ساده (Simplilearn)

مشترکین: K 883 ، فیلم ها: 2468

یادگیری ساده در واقع یک بوت کمپ آنلاین در جهان است که به افراد کمک می کند تا مهارت های لازم برای پیشرفت در اقتصاد را بدست آورند.

مرحله 3: آشنایی با کتابخانه های پایتون برای تحلیل داده



به طور کلی، پایتون بهترین روش را برای انجام هر کاری ارائه می دهد. سه کتابخانه NumPy، Pandas و Matplotlib برای تحلیل داده با پایتون مورد استفاده قرار می گیرند. در ادامه به معرفی این سه کتابخانه خواهیم پرداخت:

NumPy: کتابخانه ای که انواع عملیات ریاضی و آماری را آسان می کند. همچنین مبنای بسیاری از ویژگی های کتابخانه ی پاندا است.

pandas: یک کتابخانه پایتون که به طور خاص برای تسهیل کار با داده ها ایجاد شده است. همچنین این کتابخانه، پایه ی اصلی بسیاری از کارهای پایتون برای علوم داده است.

Matplotlib: یک کتابخانه تجسم داده است که تولید نمودارها را از داده های شما سریع و آسان می کند.

NumPy و Pandas برای کاوش و بازی با داده ها عالی هستند و Matplotlib نمودارهایی مانند آنچه در Excel یا Google Sheets مشاهده می کنید را ایجاد می کند.

ابزارهایی مورد نیاز برای تحلیل داده با پایتون

پایتون دارای یک جامعه ثروتمند متخصص است که مشتاقانه به شما در یادگیری این زبان کمک می کنند. منابعی مانند: Quora، Stack Overflow و... می تواند به شما در یادگیری برنامه نویسی پایتون کمک کند.

یکی از ابزارهایی که می تواند به شما در حوزه تحلیل داده کمک زیادی کند، ابزار Git می باشد. Git ابزار محبوبی است که به شما کمک می کند تا تغییرات ایجاد شده در کدهای خود را پیگیری کنید که این امر تصحیح اشتباهات، آزمایش و همکاری با دیگران را بسیار آسان می کند.

مرحله 4: آماده سازی نمونه کار در زمینه تحلیل داده با پایتون

به عنوان یک تحلیل گر داده، شما چند نمونه کار لازم دارید. این نمونه ها یا پروژه ها باید در زمینه های مختلف از علم داده باشند و نشان دهند که شما دارای بینشی متفاوت در زمینه ی تحلیل داده هستید. در ادامه، برخی از انواع پروژه های تحلیل داده را معرفی می کنیم:

- پروژه پاکسازی داده ها: هر پروژه ای که شامل داده های کثیف یا "غیر ساختاری" باشد که شما آن ها را تمیز و تجزیه و تحلیل می کنید، کارفرمایان بالقوه را تحت تأثیر قرار می دهد، زیرا بیشتر داده های دنیای واقعی نیاز به پاکسازی دارند.
- پروژه تجسم داده ها: ایجاد ظاهری جذاب برای داده ها که باعث شود افراد به آسانی آن را بخوانند و درک کنند، هم از جنبه ی طراحی و هم از لحاظ برنامه

نویسی شما را به چالش می کشد و اگر بتوانید آن را به درستی انجام دهید، در بیان تجزیه و تحلیل شما از داده ها، تأثیر به سزایی خواهد داشت. داشتن نمودارهای زیبا در یک پروژه باعث برجسته شدن نمونه کارها می شود.

- پروژه یادگیری ماشین: اگر شما تمایل به کار به عنوان یک دانشمند داده دارید، قطعاً به پروژه ای نیاز خواهید داشت که نمایشگرهای ML شما را به نمایش بگذارد.

تجزیه و تحلیل شما باید قابل درک و بصری ارائه شود. با استفاده از قالب هایی مانند: Jupyter Notebook ، افراد فنی می توانند کدهای شما را بخوانند و افراد غیر فنی نیز می توانند نمودارها و توضیحات کتبی شما را دنبال کنند.

نمونه کارهای شما لزوماً به یک موضوع خاص نیاز ندارند. مجموعه داده مورد علاقه خود را جست و جو کنید، سپس راهی برای ترکیب آن ها پیدا کنید. با این حال، اگر شما آرزو دارید در یک شرکت یا صنعت خاص کار کنید، به نمایش گذاشتن پروژه های مربوط به آن صنعت در نمونه کارها ایده خوبی است.

نمایش پروژه هایی از این دست به دانشمندان داده دیگر اجازه می دهد تا به طور بالقوه با شما همکاری کنند و به کارفرمایان آینده نشان می دهد که واقعاً برای یادگیری پایتون و سایر مهارت های مهم برنامه نویسی وقت صرف کرده اید.

یکی از نکات خوب در مورد داشتن نمونه کار در حوزه تحلیل داده با پایتون در رزومه این است که مهارت های شما مانند برنامه نویسی پایتون را دو برابر برجسته می کند.

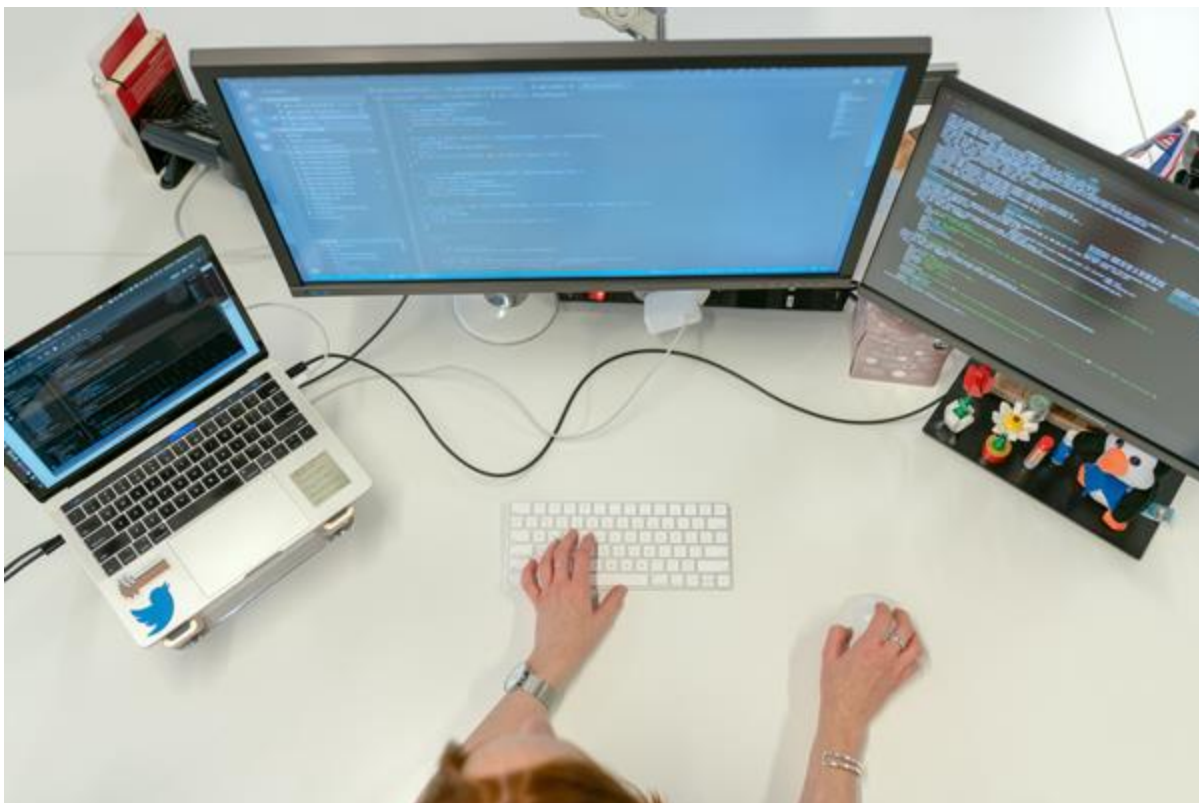
یادگیری علم آمار در سطح مبتدی و متوسط

در حالی که تحلیل داده با پایتون را یاد می‌گیرید، بهتر است که از پیش زمینه آماری نیز برخوردار شوید. درک آمار به شما این ذهنیت را می‌دهد که باید روی موارد مورد نیاز تمرکز کنید و می‌توانید راه حل‌های واقعی مناسب تری را پیدا کنید.

مرحله 5: یادگیری تکنیک‌های پیشرفته تحلیل داده با پایتون

در آخر، هدف شما این است که مهارت‌های خود را افزایش دهید. سفر شما در زمینه تحلیل داده با پایتون پر از یادگیری مداوم خواهد بود، اما دوره‌های پیشرفته‌ای وجود دارد که می‌توانید برای اطمینان از اینکه همه پایه‌ها را پوشش داده‌اید، شرکت کنید. در این مرحله، درباره‌ی موضوعاتی مانند: تحلیل رگرسیون، طبقه‌بندی، خوشه‌بندی کی و یادگیری ماشین مطالعه کنید و به یاد داشته باشید که باید به یادگیری ادامه دهید! علم داده یک زمینه رو به رشد است که صنایع مختلفی را در بر می‌گیرد و فرصت‌های بسیاری برای شما در این حوزه وجود خواهد داشت.

پاسخ به سوالات شما درباره‌ی تحلیل داده با پایتون



در ادامه به متداول ترین سوالاتی که مردم از ما در زمینه تحلیل داده با پایتون می پرسند، پاسخ می دهیم:

یادگیری پایتون چقدر طول می کشد؟

تخمین های زیادی برای مدت زمان یادگیری تحلیل داده با پایتون وجود دارد. به طور خاص برای علم داده، از سه ماه تا یک سال تمرین مداوم را تخمین می زنند.

با شرکت در دوره های **Data Scientist with python** در سماتک می توانید، سفر خود را آغاز کنید. ما مشاهده کرده ایم، افرادی که در دوره های سماتک شرکت کرده اند، سریعتر از دیگران مسیر تحلیل داده با پایتون را طی می کنند.

کجا می توانم پایتون را برای تحلیل داده یاد بگیرم؟

هزاران منبع یادگیری برای پایتون وجود دارد، اما اگر می خواهید تحلیل داده با پایتون را یاد بگیرید، بهتر است جایی را انتخاب کنید که به طور خاص در مورد علم داده آموزش می دهد.

زیرا پایتون در انواع دیگر برنامه نویسی از توسعه بازی گرفته تا برنامه های موبایلی نیز استفاده می شود. منابع عمومی یادگیری پایتون سعی می کنند کمی از همه چیز را آموزش دهند، اما این بدان معنی است که شما چیزهای زیادی را خواهید آموخت که در واقع مربوط به تحلیل داده نیستند.

اگر در این زمینه جدی هستید، بهترین راه حل یافتن بستری است که بتواند شما را برای حضور در بازار کار آماده کند. برنامه درسی ما در سماتک به نحوی است که شما را برای حضور در بازار کار تحلیل داده با پایتون آماده می کند.

آیا یادگیری پایتون برای تحلیل داده ضروری است؟

به عنوان دانشمند داده، می توانید از زبان برنامه نویسی پایتون یا R استفاده کنید. هر دو زبان نقاط قوت و ضعف خود را دارند و هر دو در صنعت به طور گسترده ای مورد استفاده قرار می گیرند. در کل تحلیل داده با پایتون محبوبیت بیشتری دارد، اما R در برخی صنایع (به ویژه در دانشگاه ها و تحقیقات) بیشتر مورد استفاده قرار می گیرد.

برای فعالیت در حوزه علوم داده، حداقل باید یکی از این دو زبان را یاد بگیرید. البته در کنار یادگیری یکی از زبان های برنامه نویسی اصلی Python یا R، باید کمی هم به SQL تسلط داشته باشید.

آیا پایتون برای علم داده بهتر از R است؟

این یک بحث همیشگی در علم داده است، اما پاسخ این سوال به آنچه که شما به دنبال آن هستید و آنچه که دوست دارید، بستگی دارد.

زبان R با در نظر گرفتن آمار و ریاضیات ساخته شده است و ویژگی هایی دارد که در تحلیل داده مورد نیاز است. همچنین یک انجمن آنلاین برای پشتیبانی از آن وجود دارد.

پایتون زبان بسیار بهتری است که می توان از آن در سایر فعالیت ها استفاده کرد، به این معنی که مهارت های پایتون شما در سایر رشته ها قابل انتقال است. همچنین تحلیل داده با پایتون کمی محبوب تر بوده و برخی معتقدند که یادگیری آن آسان تر است.

آیا آماده یادگیری تحلیل داده با پایتون هستید؟

در دوره تحلیل داده با پایتون (Data Science with Python) در سماتک، خواهیم آموخت:

چگونه داده ها را مرتب، تجزیه و تحلیل و پیاده سازی کنیم.

چگونه از کتابخانه های پایتون برای تحلیل داده استفاده کنیم.

با یادگیری ماشین و یادگیری عمیق آشنا خواهیم شد.

و ...

برای آشنایی بیشتر با دوره های Data Science with Python در سماتک، بخشی از این دوره را به صورت رایگان مشاهده کنید.