



مبانی تحلیل داده و هوش مصنوعی

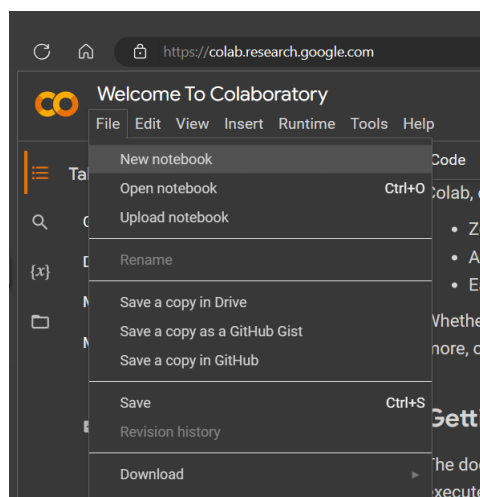
تکلیف اول

سلام بچه ها

امیدواریم حالتون خوب باشه و تا اینجا تابتون بهتون خوش گذشته باشه!

توی این Document میخوایم تکلیف اولتون رو توضیح بدیم اما همین اول بدونید که خود تکلیف خیلی حجمش کمه ولی قبل از انجامش لازمه که یسری چیز هارو بدونید و واسه همینه که یکم طولانی بنظر میرسه، پس اصلا نترسید :) .

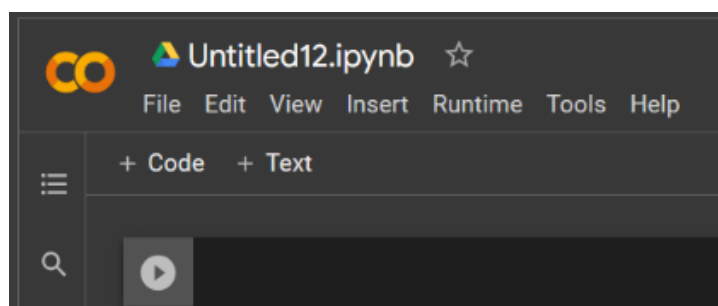
بچه ها اول از همه لازمه محیطی که باید توش کد بزنین رو بشناسید. در حال حاضر یکی از بهترین برنامه ها برای تحلیل داده DataSpell هست. اما ازونجایی که حجم برنامه تقریبا زیاده و برای نصب و کرک کردنش ممکنه به یکسری مشکلاتی بخورید (که البته قابل حله ولی زمان بره) برای همین ما تصمیم گرفتیم که بجاش محیط Google Colab رو معرفی کنیم، چون آنلاینه و تمامی کتابخونه هایی که نیاز داریم رو روی سرورش داره و اصلا لازم نیست درگیر نصب اونها بشیم! البته با این حال پیشنهاد میکنیم اگر دوست دارید خودتون DataSpell رو بریزید و یکم باهاش کار کنید. (منبعی هم خواستید برای یادگیریش حتما بگید بهمون که معرفی کنیم)



Google colab یک محیط آنلاین برای اجرای کدهای زبان پایتون هست. برای ورود بهش میتونید از این لینک استفاده کنید:

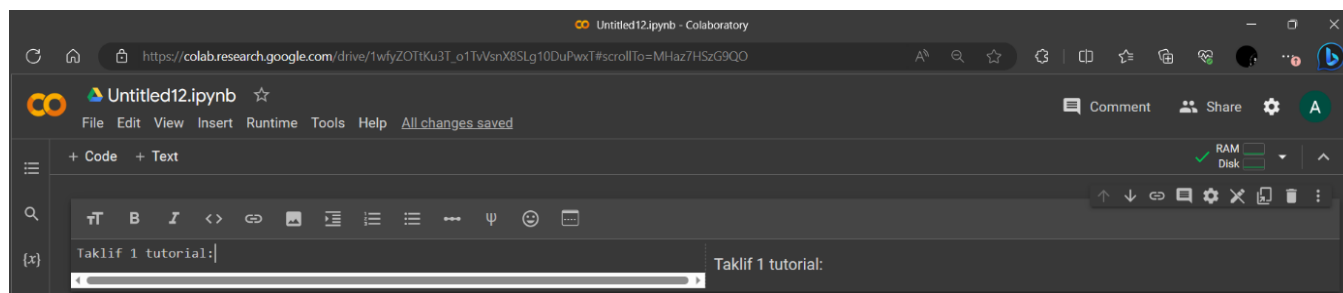
<https://colab.research.google.com>

وارد که شدید (ترجیحا با اکانت گوگل تون, اگر ندارید هم بهتره بسازید چون کلا به دردتون میخوره) بالا سمت چپ وارد File شوید و New notebook رو انتخاب کنید

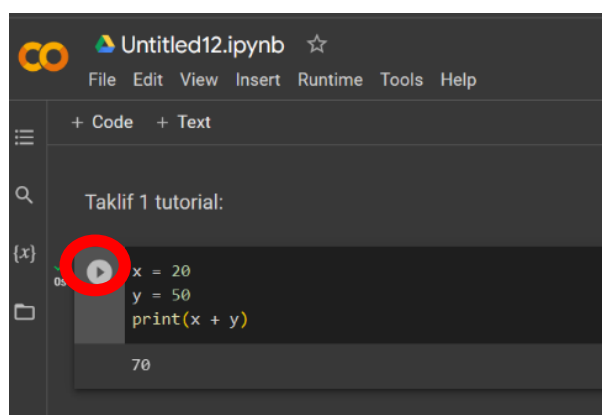
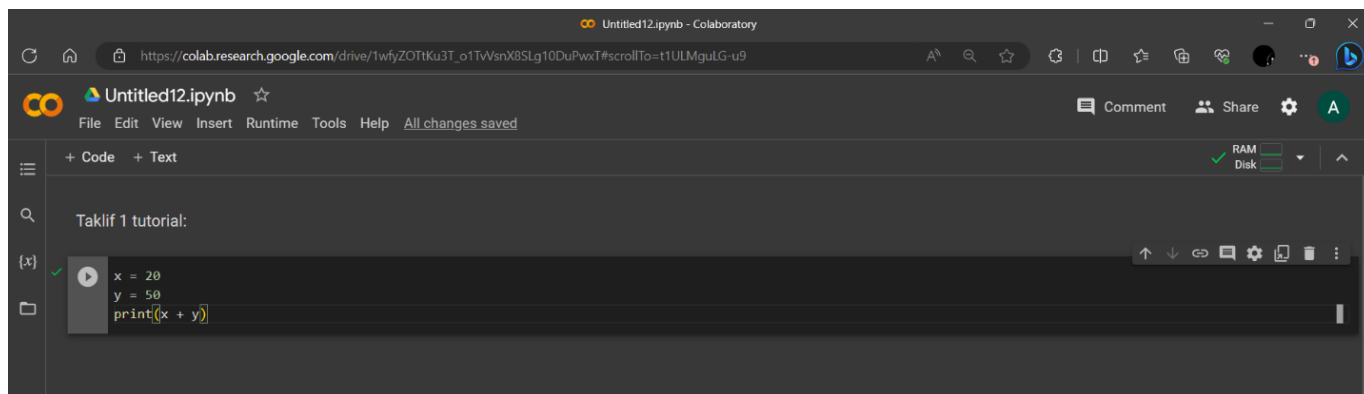


توی صفحه جدید دو گزینه +Code و +Text رو میبیند که باهاشون میتونید Cell (یا همون سلول, که اینجا منظور بخش هست) مخصوص متن یا کد اضافه کنید.

برای مثال من یک سلول متن اضافه کردم و توش متن “Taklif 1 tutorial:” رو نوشتم:



در ادامه یک سلول کد اضافه کردم و توش یک کد بسیار ساده نوشتم که x و y رو به ترتیب 50 و 20 تعریف کردم و جمع اونهارو با دستور `print()` میخوام نمایش بدم:

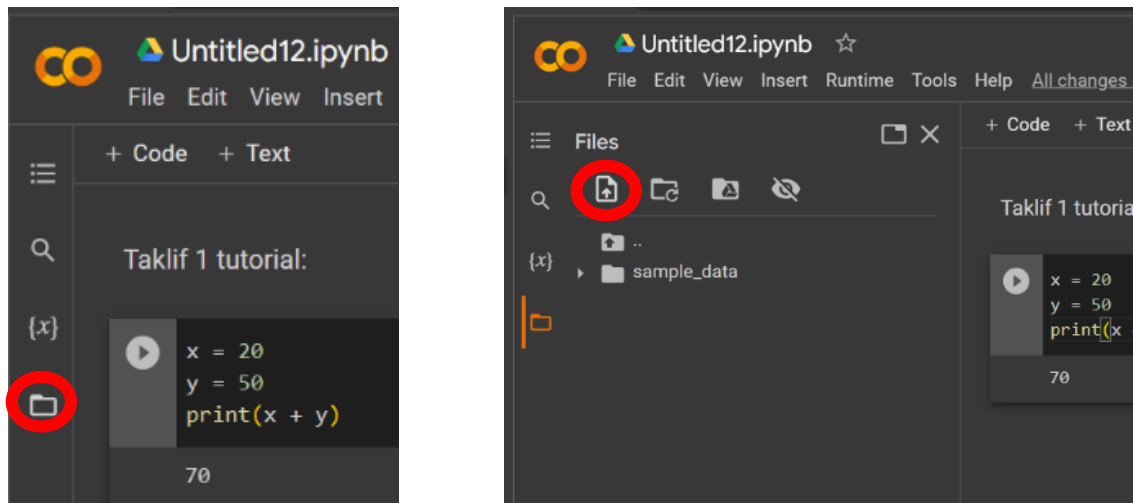


حالا برای اجرای کد کافیه روی دایره فلش دار کنار سلول کلیک کنم تا کد Run بشه و خروجی مد نظر نمایش داده بشه:

همونطور که میبینید خروجی جمع 20 و 50 رو در زیر سلول 70 نمایش داده که درسته.

در توضیح اینکه سلول متن به چه دردی میخوره لازمه بدونید که فرم درست و قابل قبول کد زدن در Notebook ها این جوری هست که شما بالای هر سلول کد باید یک سلول متن داشته باشید که توضیحات کاری که توی اون سلول کد قرار هست انجام بدید رو به صورت خیلی خلاصه بنویسید و اینجوری نظم خیلی زیادی به Notebook تون میدید و خوندنش برای خودتون هم راحت تر میشه. پس همیشه در همه تکالیفتون این قانون رو رعایت کنید!

همچنین ازونجایی که ما قرار هست با فایل های داده کار کنیم، باید فایل دادمون رو هم در این محیط مجازی آپلود کنیم تا بتونیم ازش استفاده کنیم و کارهایی که میخوایم رو روش پیاده کنیم. برای این کار در سمت چپ روی بخش Files کلیک میکنیم و گزینه Upload to session storage رو انتخاب میکنیم تا بتونیم فایل مدنظرمون رو آپلود کنیم.



عموما فایل های دادمون فرمت CSV. دارن که درباره این فرمت سر کلاس توضیح میدیم. وقتی فایل داده رو آپلود کردیم باید توی کد پایتونمون اون رو به یک Data Frame تبدیل کنیم تا بتونیم محاسباتی که میخوایم رو روش انجام بدیم. (خیلی درگیر این نشید که از کجا اومد، صرفا بپذیرید که توی پایتون همیشه مستقیم رو فایل داده کاری کرد و حتما باید قبلش به یک دیتا فریم تبدیل بشه)

برای اینکار باید از کتابخونه Pandas کمک بگیریم و اینکه کتابخونه ها چی هستن و چه کمکی به ما میکنن رو هم جلسه آینده کامل براتون توضیح میدیم با مثال ولی الان فقط لازمه کد مربوط بهش رو یاد بگیرید و بتونید یکبار بنویسیدش.

برای استفاده از هر کتابخانه ای در زبان پایتون، باید با یک دستور اون رو در ابتدای کدمون الصاق کنیم تا کامپیوتر بفهمه از کدوم کتابخانه ها قرار هست استفاده کنیم. دستور مربوط به این کار رو میتونید ببینید:

```
import pandas
```

بعد از اینکه کتابخانه مورد نظر رو import کردیم حالا باید دادمون رو به Data Frame تبدیل کنیم. کد مربوط به این کار رو در زیر میتونید ببینید:

```
import pandas  
  
DataFrame = pandas.read_csv("File Name")
```

درون کوتیشن ها (همون علامت ") باید اسم فایل دادتون رو که آپلود کردین بنویسید، مثلا اگر فایل داده Data.csv هست باید بنویسیم:

```
DataFrame = pandas.read_csv("Data.csv")
```

حالا که اینکارو کردیم میتونیم از دیتافریم توضیحاتی مثل میانگین، میانه، چارک دوم تا سوم و انحراف معیار (که سر کلاس کار کردیم) رو بگیریم.

برای این کار میتونیم تک تک فرمول هرکدوم رو به کد پایتون تبدیل کنیم و دستی هرکدوم رو برای داده محاسبه کنیم اما، اصلا لازم نیست این کار سخت و زمان بر رو انجام بدیم!!

چون کتابخانه Pandas خودش دستور describe (توضیح) داره که تمام اینهارو برای ما محاسبه میکنه و به عنوان خروجی میده. کد مربوط بهش رو اینجوری مینویسیم:

```
import pandas

DataFrame = pandas.read_csv("Data.csv")
DataFrame.describe()
```

و وقتی کد رو اجرا کنیم به عنوان خروجی یک جدول این شکلی به ما میده:

```
import pandas

DataFrame = pandas.read_csv("Data.csv")
DataFrame.describe()
```

	ExamFieldID	IsOnline	NofParticipants
count	17.000000	17.000000	17.000000
mean	1800.588235	0.529412	12477.058824
std	35.323609	0.514496	4717.008698
min	1739.000000	0.000000	6489.000000
25%	1775.000000	0.000000	8414.000000
50%	1805.000000	1.000000	11125.000000
75%	1829.000000	1.000000	17124.000000
max	1853.000000	1.000000	19327.000000

همونطور که تو جدول هم میتونید ببینید هر ردیف به ترتیب نشون دهنده این ها هستند:

1. تعداد
2. میانگین
3. انحراف معیار (Standard Deviation – STD)
4. مینیمم (کمترین مقدار داده)
5. چارک اول
6. چارک دوم (میانه)
7. چارک سوم
8. ماکسیمم (بیشترین مقدار داده)

همچنین میتونید ببینید که این اعداد 3 سری محاسبه شدن و در 3 ستون قرار گرفتن که بالای هر ستون یک اسم نوشته شده. اون اسم ها، اسم ستون ها در فایل داده من هستن. برای مثال در فایلی که من آپلود کردم سه ستون به نام های ExamFieldID, IsOnline, NofParticipants وجود داشت، در نتیجه دستور describe همه معیار هارو برای هر ستون جداگانه محاسبه کرد و اون هارو در یک جدول سه ستونه ارائه کرد.

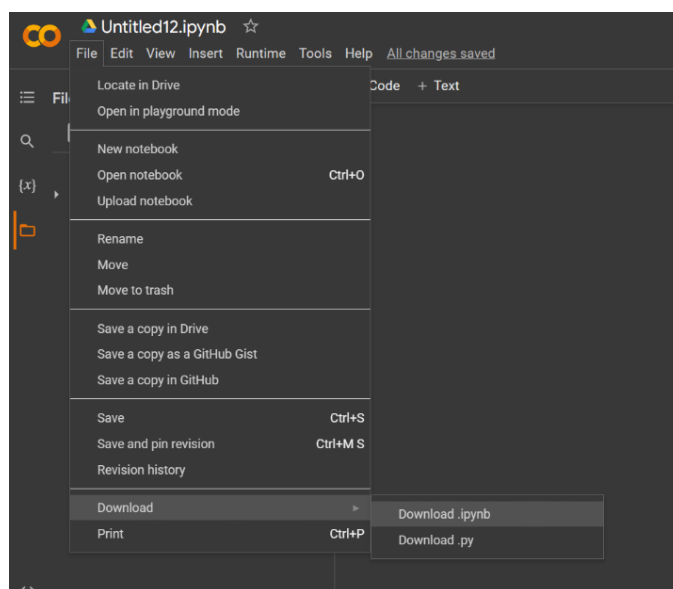
و اما تکلیفتون:

ما یک فایل داده بهتون میدیم که مربوط به افرادی هست که سوار کشتی تایتانیک شده بودند! اول از همه خودتون فایل رو با Excel باز کنید و کمی بالا پایینش کنید تا بفهمید شکل داتتون چجوریه، چه اطلاعاتی درس هست و هر ستون نمایشگر چه چیزی هست.

بعد با استفاده از محیط Google colab دیتا رو به یک DataFrame تبدیل کنید و جدول مربوط به معیار های اصلی داده رو ازش خروجی بگیرید.

برای تحویل تکلیف بعد از اینکه کدتون با موفقیت اجرا شد، فایلتون رو با فرمت `ipynb` دانلود

کنید:



اسمش رو به نام و نام خانوادگی خودتون تغییر بدید و درلینکی که توی گروه میفرستیم آپلود کنید.

حتما حتما فرمت فایل رو `ipynb` انتخاب کنید نه `py`.

مهلت تحویل تا یکشنبه شب هست پس حواستون باشه دیر نکنید.

موفق باشید