



تمرین سوم کارگاه پایتون

موعد تحویل: شنبه - ۱۵ اردیبهشت ۱۳۹۷

تمرین را انفرادی انجام دهید

این تمرین ۲ نمره از نمره پایانی شما را تشکیل می‌دهد

نوع: کد + گزارش کار

لطفا قبل از انجام تمارین اسلایدهای جلسه ششم و سه نوت بوک جلسه ۷ را بخوانید.

سوال (۱) ۰,۲۵ نمره

سوالاتی که در نوت بوک

7-2_Intro_Data_Analysis.ipynb

با قرمز مشخص شده را بنویسید و نوت بوک تکمیل شده را ارسال کنید.

سوال (۲) ۰,۲۵ نمره

سوالاتی که در نوت بوک

7-3_Data_Wrangling.ipynb

با قرمز مشخص شده را بنویسید و نوت بوک تکمیل شده را ارسال کنید.

سوال (۳) ۱,۵ نمره

مقدمه تمرین :

تمرین های چند سری آینده از چالش دو سال پیش سایت Kaggle انتخاب شده است.

Titanic: Machine Learning from Disaster

<https://www.kaggle.com/c/titanic>

در این چالش ویژگی هایی از مسافران کشتی تایتانیک استخراج شده و عاقبت هر مسافر در ستونی به نام survival که نشان دهنده نجات یا عدم نجات مسافر بوده آمده است.

هدف این است که بتوانیم بر اساس این ویژگی ها و اطلاعات بتوانیم برای سایر مسافرینی که از سرنوشت آنها خبر نداریم پیش بینی درستی از عاقبت آنها بیابیم.

در گام نخست باید دیتا را برای مراحل بعدی آماده کنیم.

توجه: دیتاست برای تمرین بیشتر کمی تغییرات جزئی شامل حذف عنوان و تغییر نوع سن داشته که شما در فرآیند پیش پردازش آنها را باید درست کنید.

دریافت مجموعه داده:

مجموعه داده:

http://pylab.akhavanpour.ir/dataset/titanic_train.csv

عنوان ستون ها:

http://pylab.akhavanpour.ir/dataset/titanic_headers.txt

توضیحات بیشتر:

<http://pylab.akhavanpour.ir/dataset/titanic-kaggle.docx>

توضیح داده ها:

مجموعه داده دارای عناوین ستون نیست!

برخی از داده ها شامل missing value است. باید اصلاح شود.

نوع داده سن اشتباها object است و نیاز دارد به اعشاری تغییر یابد!

در داده های جنسیت برای آقا male و برای خانم هم f و هم female ثبت شده و نیاز به اصلاح دارد!

توضیح تمرین:

در ابتدا داده ها را بر اساس توضیح بالا و مطالبی که در کلاس خواندید تمیز کنید. سپس گزارشی از مواردی که برای تمیز کردن داده انجام داده اید را به همراه توضیحات داده شامل توضیح فارسی هر فیلد و تحلیل ساده از قبیل تحلیل های عددی شامل میانگین و تحلیل های غیر عددی شامل بیشتر مقدار و ... ارائه دهید.

فایل گزارش حداقل یک صفحه خواهد بود.

در انتها گزارش خود را در قالب یک فایل pdf و فایل zip کدهایتان را ارسال خواهید کرد.