# پروژه ی ABC

شرکت تاکسیرانی ABC آژانسی است که مسئول صدور مجوز و تنظیم تاکسی ها و وسایل نقلیه اجاره ای شهر نیویورک است. این آژانس میخواهد یک مدل رگرسیونی ایجاد کند که به تخمین کرایه تاکسی قبل از سوار شدن کمک کند.

داده های ABC از بیش از 200000 دارنده مجوز تاکسی بدست می آید که تقریباً یک میلیون سفر در روز انجام می دهند.

پروژه ی پایانی سناریو ABC، از دو بخش EDA و Regression تشکیل شده است.

## داده های ABC

داده های این سناریو در فایل ABC\_Taxi\_Data.csv قرار دارد و شامل **408,294 rows می باشد. هر سطر بیانگر انجام یک سفر است**. داده ها شامل 18 ویژگی به صورت زیر میباشند:

|  |  |
| --- | --- |
| **Column name** | **Description** |
| ID | Trip identification number |
| VendorID | A code indicating the TPEP provider that provided the record.  **1= Creative Mobile Technologies, LLC;**  **2= VeriFone Inc.** |
| tpep\_pickup\_datetime | The date and time when the meter was engaged. |
| tpep\_dropoff\_datetime | The date and time when the meter was disengaged. |
| Passenger\_count | The number of passengers in the vehicle.  This is a driver-entered value. |
| Trip\_distance | The elapsed trip distance in miles reported by the taximeter. |
| PULocationID | TLC Taxi Zone in which the taximeter was engaged |
| DOLocationID | TLC Taxi Zone in which the taximeter was disengaged |
| RateCodeID | The final rate code in effect at the end of the trip.  **1= Standard rate**  **2=JFK**  **3=Newark**  **4=Nassau or Westchester**  **5=Negotiated fare**  **6=Group ride** |
| Store\_and\_fwd\_flag | This flag indicates whether the trip record was held in vehicle memory before being sent to the vendor, aka “store and forward,” because the vehicle did not have a connection to the server.  **Y= store and forward trip**  **N= not a store and forward trip** |
| Payment\_type | A numeric code signifying how the passenger paid for the trip.  **1= Credit card**  **2= Cash**  **3= No charge**  **4= Dispute**  **5= Unknown**  **6= Voided trip** |
| Fare\_amount | The time-and-distance fare calculated by the meter. |
| Extra | Miscellaneous extras and surcharges. Currently, this only includes the $0.50 and $1 rush hour and overnight charges. |
| MTA\_tax | $0.50 MTA tax that is automatically triggered based on the metered rate in use. |
| Improvement\_surcharge | $0.30 improvement surcharge assessed trips at the flag drop. The  improvement surcharge began being levied in 2015. |
| Tip\_amount | Tip amount – This field is automatically populated for credit card tips. Cash tips are not included. |
| Tolls\_amount | Total amount of all tolls paid in trip. |
| Total\_amount | The total amount charged to passengers. Does not include cash tips. |

## بخش اول EDA

در این بخش، جهت آشنایی با داده ها از مراحل EDA استفاده می شود. همچنین از power bi جهت نمایش نتایج به ذینفعان غیرفنی استفاده می شود.

داکیومنت های زیر در انتهای این بخش میبایست ارائه شود:

* تکمیل سندCourse 2 PACE Strategy
* انجام کد نویسی در Jupyter notebook: این فایل به کمک یک فریم ورک راهنما، به شما کمک می کند تا کدهای مورد نیاز EDA و نمایش دادها (visualization) را انجام دهید. همچنین از طریق پاسخ به سوالاتی که در این فایل قرار گرفته و سند PACE، می توانید فایل executive summary را نیز تکمیل نمایید.
* ساخت نمایش داد ها در Jupyter notebook و power bi (استفاده از power bi اختیاری می باشد و شامل نمره ی مثبت است)
* تکمیل فایل executive summary. این فایل که به صورت یک صفحه ای میباشد به شما اجازه می دهد تا به صورت سطح بالا، نتایج و مطالب انجام شده را برای ذینفعان ارائه دهید. از جمله این مطالب شامل خلاصه ای از نتایج EDA و راه حل برای حل مشکل داده های پرت می باشد. (یکی از مدل های فایل executive summary را انتخاب و تکمیل نمایید)

فایل های ذکر شده را به صورت فایل زیپ در سامانه ی ایلرن آپلود نمایید.