2026

StudyBuild - Project 01

Data Cleaning on an E-commerce Dataset
Project Guide

Prepared byAtefe Asadi |

پروژه اول Study&Build

---

۱ .معرفی دیتاست

این دیتاست مربوط به اطلاعات فروش و سفارش‌های یک فروشگاه اینترنتی است و شامل اطلاعاتی مانند:

- شناسه سفارش(Order ID)
- شناسه مشتری(Customer ID)
- محصول(Product)
- تعداد خرید(Quantity)
- قیمت(Price)
- تخفیف(Discount)
- شهر(City)
- تاریخ سفارش(Order Date)
- روش پرداخت(Payment Method)

و سایر اطلاعات مرتبط با سفارش‌ها است.

این داده یک **دیتاست خام (Raw Data)** محسوب می‌شود و مانند بسیاری از داده‌های واقعی، قبل از تحلیل نیاز به بررسی و پاک‌سازی دارد.

---

۲ .صورت مسئله

فرض کنید به‌عنوان یک **Data Analyst** در یک فروشگاه اینترنتی استخدام شده‌اید و این فایل خام در اختیار شما قرار گرفته است.

مدیر مجموعه قصد دارد از این داده برای تحلیل فروش، رفتار مشتریان و تصمیم‌گیری‌های تجاری استفاده کند؛ اما قبل از هر تحلیلی باید مطمئن شویم داده‌ها دقیق، منظم و قابل اعتماد هستند.

وظیفه شما این است که کیفیت داده را بررسی کرده، مشکلات آن را شناسایی و اصلاح کنید و در نهایت یک نسخه تمیز و آماده تحلیل از دیتاست ارائه دهید.

---

۳ .مشکلات احتمالی موجود در دیتاست

ممکن است با موارد زیر روبه‌رو شوید:

- مقادیر گمشده(Missing Values)
- رکوردهای تکراری(Duplicate Records)
- نوع داده اشتباه(Incorrect Data Types)
- قالب‌های متفاوت تاریخ(Inconsistent Date Formats)
- مقادیر غیرمنطقی(Invalid Values)

- داده‌های پرت(Outliers)
- فاصله‌های اضافی در متن‌ها
- نام‌های ناهماهنگ یا با نگارش متفاوت
- خطاهای تایپی
- ناسازگاری بین ستون‌ها

ممکن است علاوه بر موارد بالا، مشکلات دیگری نیز در داده وجود داشته باشد. اگر مورد جدیدی پیدا کردید، آن را مستند کنید.

---

۴ .هدف پروژه

هدف این پروژه فقط حذف داده‌های تکراری یا اصلاح چند مقدار خالی نیست.

هدف اصلی این است که یاد بگیرید چگونه یک دیتاست خام را به یک دیتاست استاندارد، تمیز و آماده تحلیل تبدیل کنید؛ دقیقاً همان کاری که یک تحلیلگر داده در پروژه‌های واقعی انجام می‌دهد.

پس از پایان این پروژه، این دیتاست در پروژه‌های بعدی نیز مورد استفاده قرار خواهد گرفت.

---

۵ .در ادامه این دیتاست چه استفاده‌ای خواهد شد؟

بعد از پاک‌سازی داده، روی همین دیتاست پروژه‌های بعدی انجام می‌شود؛ از جمله:

- تحلیل اکتشافی داده(EDA)
- مصورسازی داده‌ها(Data Visualization)
- تحلیل رفتار مشتریان
- تحلیل فروش
- طراحی داشبورد
- پروژه‌های یادگیری ماشین

بنابراین کیفیت Data Cleaning شما مستقیماً روی کیفیت پروژه‌های بعدی تأثیر خواهد گذاشت.

---

۶ .سؤال‌هایی که باید بتوانیم به آن‌ها پاسخ دهیم

پس از پاک‌سازی داده انتظار داریم بتوانیم به سؤال‌هایی مانند موارد زیر پاسخ دهیم:

- کدام محصولات بیشترین فروش را داشته‌اند؟
- کدام شهرها بیشترین درآمد را ایجاد کرده‌اند؟
- کدام مشتریان بیشترین خرید را انجام داده‌اند؟
- میانگین مبلغ هر سفارش چقدر است؟
- کدام روش پرداخت بیشترین استفاده را داشته است؟
- تخفیف چه تأثیری بر میزان فروش داشته است؟

- فروش در چه بازه‌های زمانی بیشتر بوده است؟
- آیا داده‌ای وجود دارد که باعث نتیجه‌گیری اشتباه شود؟

اگر بتوانید سؤال‌ها یا تحلیل‌های بیشتری از دیتاست استخراج کنید، امتیاز مثبت خواهد داشت.

---

۷. وظایف شما

برای این پروژه باید مراحل زیر را انجام دهید:

1. بررسی کامل دیتاست
2. شناسایی مشکلات موجود
3. تصمیم‌گیری درباره نحوه اصلاح هر مشکل
4. پاک‌سازی داده
5. مستندسازی تمام تغییرات
6. تولید نسخه نهایی دیتاست

---

۸. خروجی‌های موردنیاز

حداقل خروجی پروژه شامل دو فایل است:

۱. دیتاست پاک‌سازی‌شده

cleaned_dataset.xlsx

۲. فایل README

README.md

در فایل README موارد زیر را توضیح دهید:

- چه مشکلاتی در داده پیدا کردید؟
- چه تغییراتی انجام دادید؟
- چرا آن تغییر را انجام دادید؟
- چگونه Missing Value ها را مدیریت کردید؟
- چگونه Duplicate ها را بررسی کردید؟
- آیا نوع داده ستونی را تغییر دادید؟
- آیا داده پرت یا مقدار غیرمنطقی پیدا کردید؟
- از چه ابزار یا کتابخانه‌هایی استفاده کردید؟
- نسخه نهایی چه تفاوتی با نسخه اولیه دارد؟

فایل‌های اختیاری

اگر پروژه را با Python یا R انجام دادید، فایل کد یا Notebook را نیز بارگذاری کنید.

data_cleaning.ipynb

یا

data_cleaning.R

همچنین اگر تمایل داشتید، می‌توانید یک گزارش کوتاه شامل تحلیل، نمودارها و نتایج خود تهیه کنید.

analysis_report.pdf

این فایل اختیاری است، اما برای ساخت پورتفولیو بسیار ارزشمند خواهد بود.

---

۹. ساختار پوشه پروژه

پیشنهاد می‌شود فایل‌ها را با ساختار زیر آماده کنید:

```
project-01-data-cleaning/
├── data/
│   └── cleaned_dataset.xlsx
├── notebook/
│   └── data_cleaning.ipynb
├── report/
│   └── analysis_report.pdf
└── README.md
```

اگر فقط با Excel کار کرده‌اید، وجود پوشه **notebook** الزامی نیست.

اگر گزارش PDF تهیه نکرده‌اید، پوشه **report** نیز لازم نیست.

---

۱۰. شیوه نام‌گذاری فایل‌ها

برای حفظ نظم Repository ، لطفاً از الگوی زیر استفاده کنید.

نام پوشه پروژه

project-01-data-cleaning-githubusername

مثال:

project-01-data-cleaning-atefeasadi

نام فایل دیتاست

cleaned_dataset_githubusername.xlsx

نامNotebook

data_cleaning_githubusername.ipynb

نام گزارش

analysis_report_githubusername.pdf

نام فایل مستندات

README.md

لطفاً از نام‌هایی مانند موارد زیر استفاده نکنید:

- final
- final2
- last
- new
- نسخه جدید
- فایل نهایی

---

۱۱ .نحوه تحویل پروژه

همه اعضا باید ابتدا **Google Form**را تکمیل کرده و نام کاربری GitHub خود را وارد کنند.

بدون GitHub Username امکان اضافه شدن به GitHub Organization وجود ندارد.

پس از اضافه شدن به:GitHub

1. وارد Repository پروژه شوید.
2. پوشه مخصوص خودتان را ایجاد کنید.
3. فایل‌های پروژه را داخل پوشه بارگذاری کنید.
4. تغییرات را Commit کنید.
5. در صورت نیاز Pull Request ارسال کنید.

آموزش کامل GitHub و نحوه بارگذاری پروژه در گروه منتشر خواهد شد.

---

۱۲ .اگر انجام پروژه را بلد نبودید

نگران نباشید.

هدف این پروژه یادگیری است، نه کامل بودن.

پیشنهاد می‌کنیم مسیر زیر را دنبال کنید:

1. ابتدا دیتاست را بررسی کنید.
2. منابع معرفی‌شده را مطالعه کنید.
3. خودتان برای حل مسئله تلاش کنید.
4. اگر به مشکل برخوردید، سؤال خود را همراه با توضیح کارهایی که انجام داده‌اید در گروه مطرح کنید.
5. می‌توانید پروژه را با **Excel**، **Python** یا **R** انجام دهید.

مهم‌ترین نکته این است که تصمیم‌های شما منطقی، مستند و قابل توضیح باشند؛ نه اینکه حتماً از روش‌های پیچیده استفاده کنید.

© 2026 Study&Build