



# پروژه مبانی داده کاوی

بررسی رفتار کاربران در یک بانک جهت افتتاح حساب

امیررضا حسینی - پاشا احمدی

بهار ۱۴۰۲

دانشگاه صنعتی اصفهان دانشکده برق و کامپیوتر

# فهرست مطالب

## ۱- مقدمه

- مشکل چیست؟
- سوال داده کاوی که برای حل مشکل مطرح شده است چیست؟
- داده های جمع آوری شده
- نحوه ارزیابی

## ۲- آزمایش ها

- اکتشافات داده ای و لیست حقایق کشف شده در این مرحله
- اکتشافات مدل سازی و لیست حقایق کشف شده در این مرحله

## ۳- نتایج

- نتیجه اجرای روش نهایی و لیست حقایق کشف شده در این مرحله
- لیست نتایج نهایی

## ۴- پیشنهادات

- پیشنهادات برای نوآوری و بهبود

## ۵- مراجع

- لیست مراجع

# مقدمه

---

- کمپین‌های تبلیغاتی سالیانه
- سپرده‌های کوتاه مدت و بلند مدت
- مشتری خوب؟ مشتری بد؟
- جذب مشتری و سوددهی
- پیشنهادات متنوع



# بررسی رفتار کاربران در یک بانک جهت افتتاح حساب

- معرفی حوزه مسئله
- سوال داده کاوی برای حل مشکل



# بیان مختصر ایده تیم

- حذف مقادیر Missing و شناسایی داده های پرت و حذف آنها
- نرمال سازی و استاندارد سازی ویژگی ها
- سبب بندی و reclassify
- در نظر گرفتن ویژگی های با بیشترین همبستگی با ویژگی هدف برای تحلیل روابط
- استخراج ویژگی های جدید با همبستگی بالا
- حذف ویژگی هایی با همبستگی کم با متغیر هدف و جایگزینی آن با ویژگی های جدید استخراج شده



داده مشتریان

---

# مارکتینگ موسسه مالی پرتقال

داده‌های مربوط به مشتریان	سن	شغل	وضعیت تاهل	موجودی
داده‌های تماس کمپین قبلی	تعداد تماس	روز هفته	مدت زمان تماس	ماه
ویژگی هدف	آیا مشتری حساب کوتاه مدت باز کرده یا نه؟			

# مشخصات دیتاست و ویژگی های آن

داده‌های مربوط به مشتریان شامل:

- سن: (عددی)
- شغل: نوع شغل (دسته‌ای: 'ادمین', 'کارگر', 'کارآفرین', 'خانه‌دار', 'مدیر', 'بازنشسته', 'خود اشتغال', 'خدمات', 'دانشجو', 'تکنسین', 'بیکار', 'نامشخص')
- وضعیت تأهل: وضعیت تأهل (دسته‌ای: 'طلاق گرفته', 'متأهل', 'مجرد', 'نامشخص'),
- تحصیلات: (دسته‌ای: سیکل، دیپلم، عالی، و نامشخص)
- قرض الحسنه: آیا شخص حساب قرض الحسنه دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- وام مسکن: آیا شخص وام مسکن دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- وام شخصی: آیا شخص وام شخصی دارد؟ (دسته‌ای: 'خیر', 'بله', 'نامشخص')
- موجودی: کل موجودی حساب شخص. (یک متغیر عددی که می‌تواند برحسب بدهکاری یا طلبکاری منفی یا مثبت باشد).



# مشخصات دیتاست و ویژگی های آن

داده‌های مربوط به آخرین تماس با هر مشتری در کمپین فعلی:

- نوع بستر ارتباط تماس (دسته‌ای: 'گوشی موبایل', 'تلفن ثابت')
  - آخرین ماه تماس در سال (دسته‌ای: 'ژانویه', 'فوریه', 'مارس', ..., 'نوامبر', 'دسامبر')
  - آخرین روز هفته تماس (دسته‌ای: 'دوشنبه', 'سه‌شنبه', ..., 'چهارشنبه', 'پنج‌شنبه', 'جمعه')
  - مدت‌زمان آخرین تماس گرفته شده با مشتری بر حسب ثانیه (عددی).
  - تعداد تماس‌های انجام شده در طول این کمپین و برای این مشتری (عددی، شامل آخرین تماس)
  - تعداد روزهایی که پس از تماس آخرین با مشتری از کمپین قبلی گذشته است. (عددی؛ عدد ۹۹۹ به معنای عدم تماس قبلی با مشتری است)
  - تعداد تماس‌های انجام شده قبل از این کمپین و برای این مشتری (عددی)
  - نتیجه کمپین بازاریابی قبلی (دسته‌ای: 'شکست', 'ناموجود', 'موفقیت')
- برچسب خروجی (Y) شامل:
- آیا مشتری سپرده کوتاه‌مدت باز کرده است؟ (باینری شامل بله و خیر)



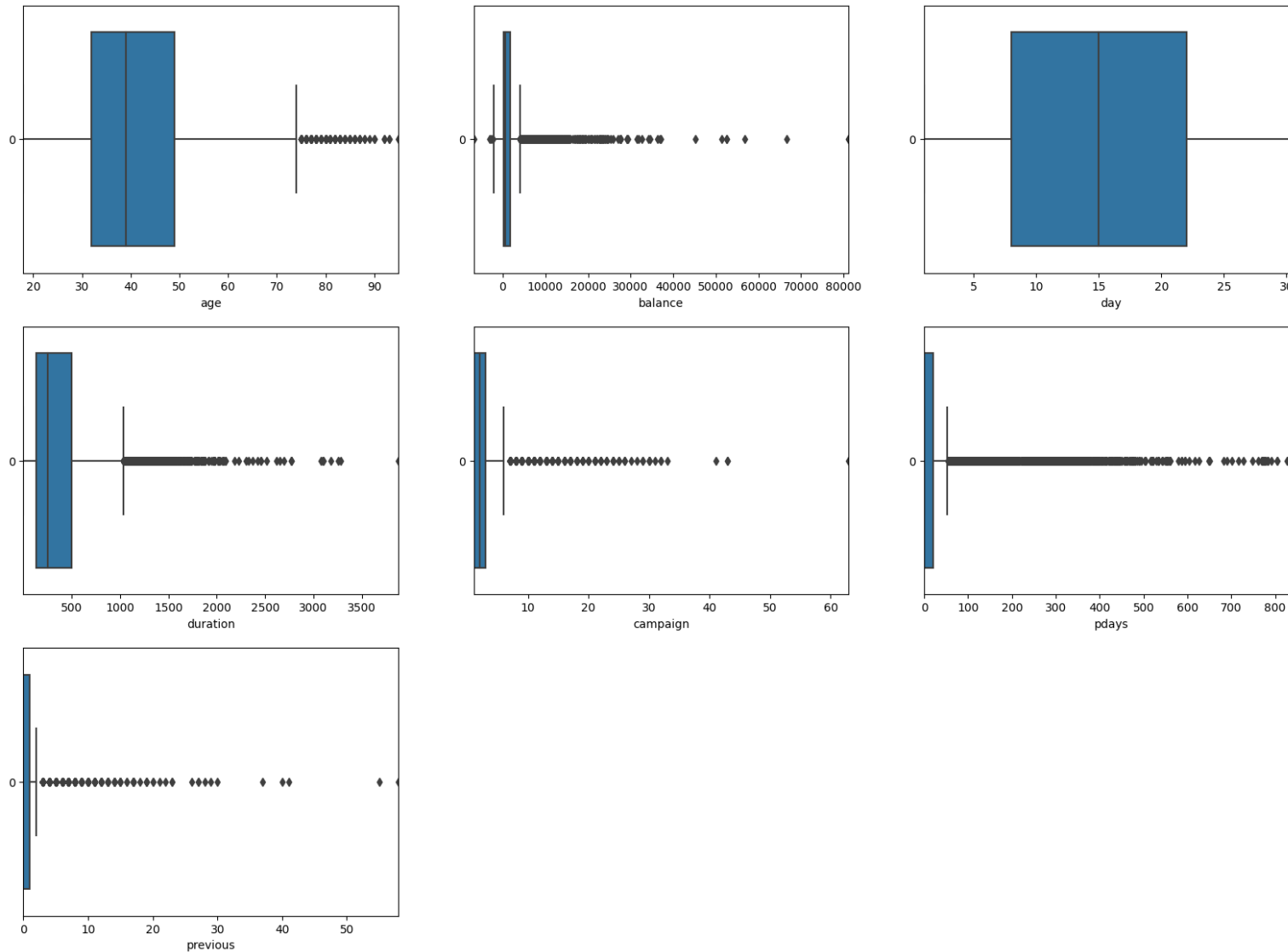
# آزمایش‌ها

---

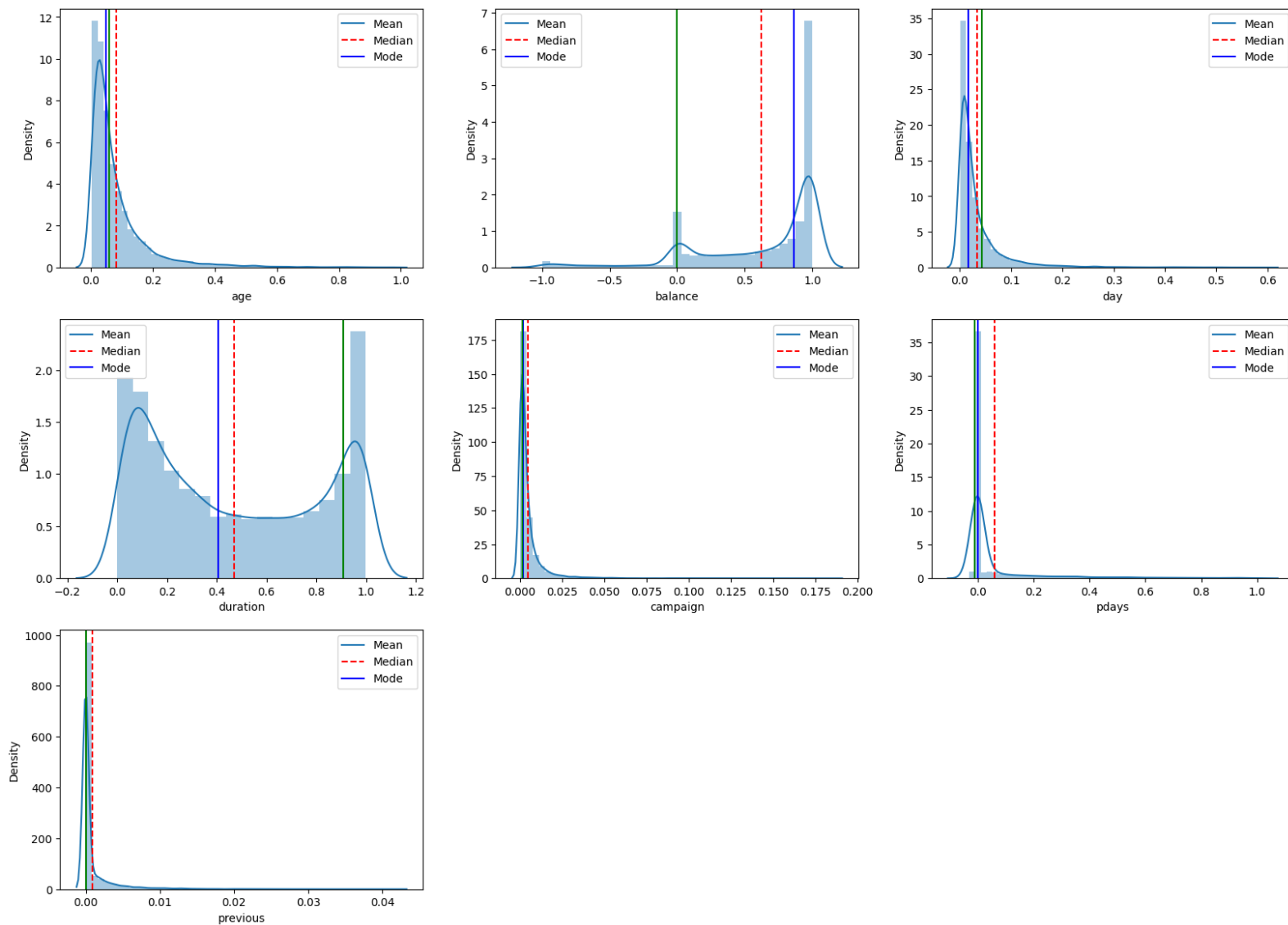




# اکتشافات داده‌ای



# اکتشافات داده‌ای



# اکتشافات داده‌ای

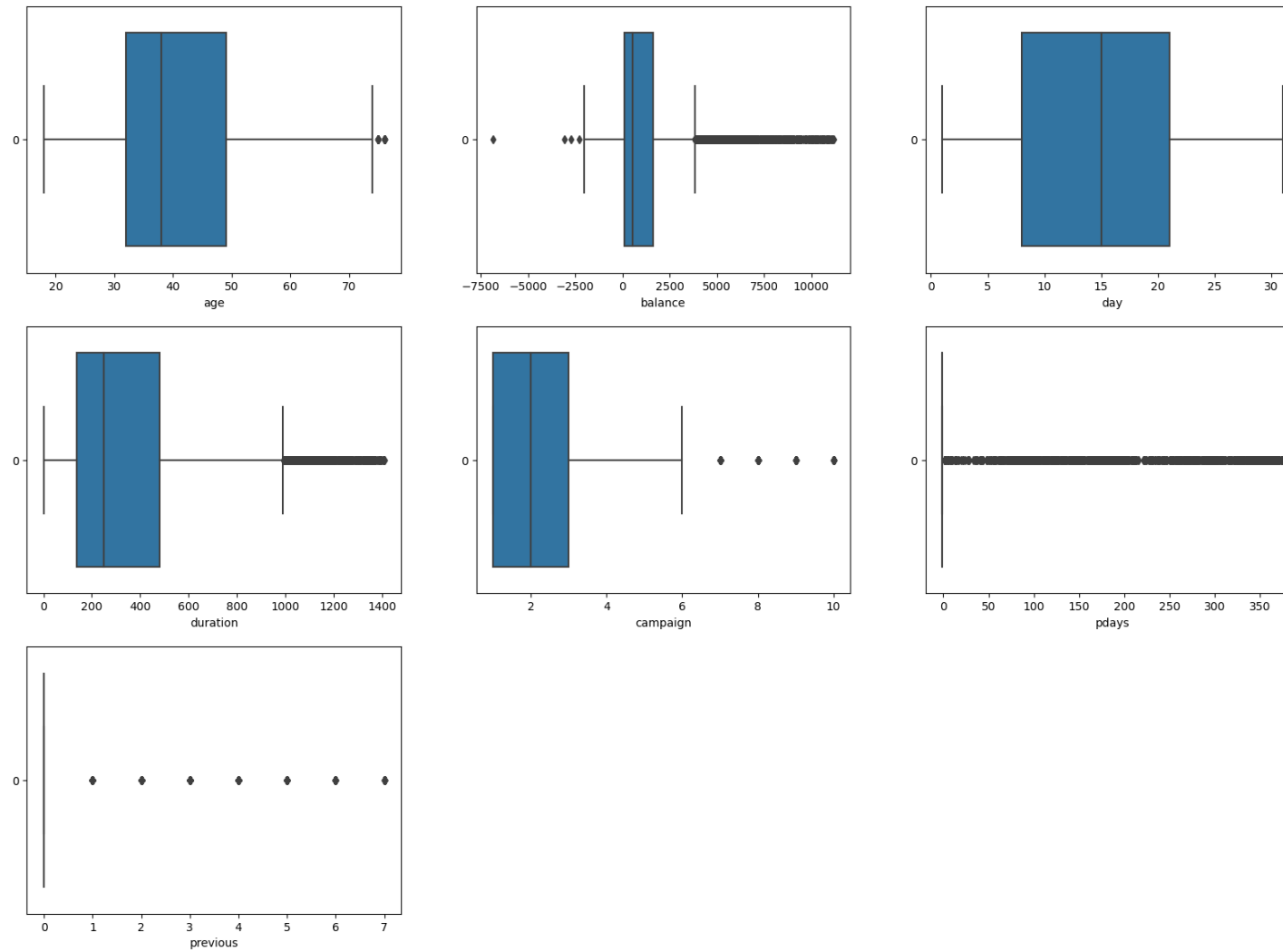
---

حذف داده‌ها پرت

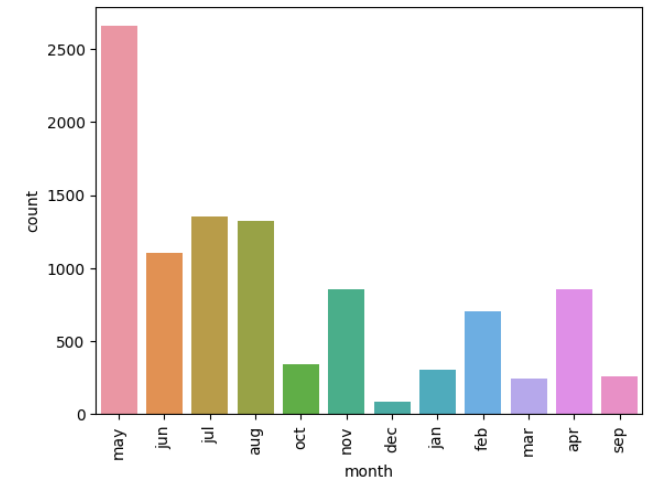
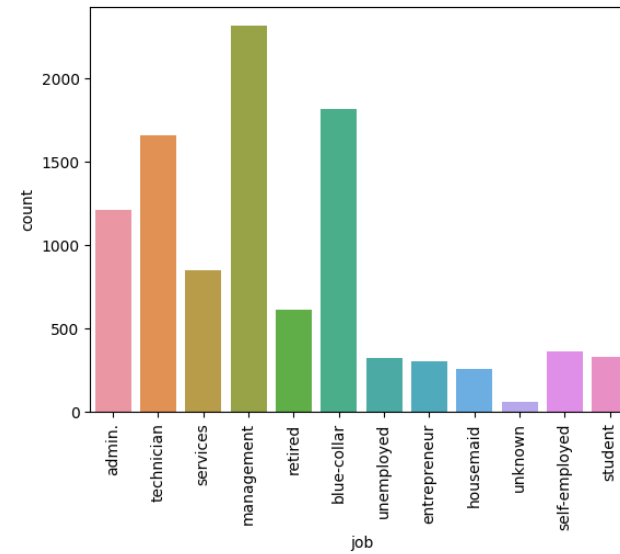
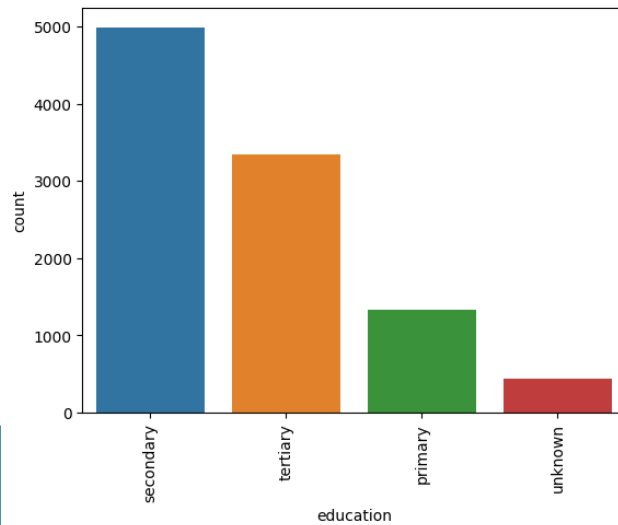
Z-score

IQR

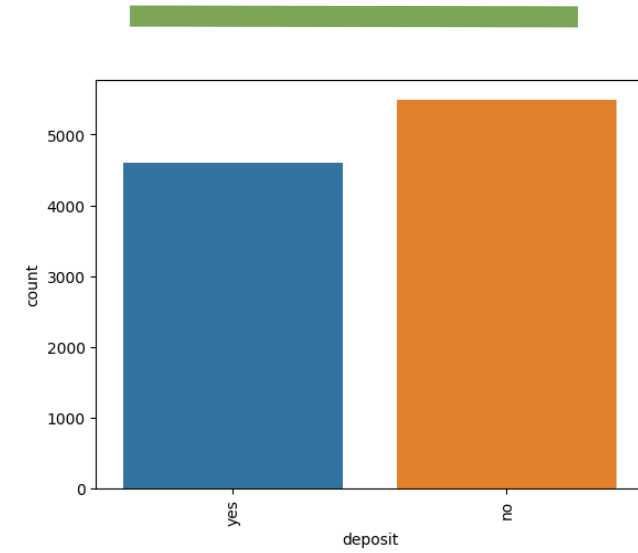
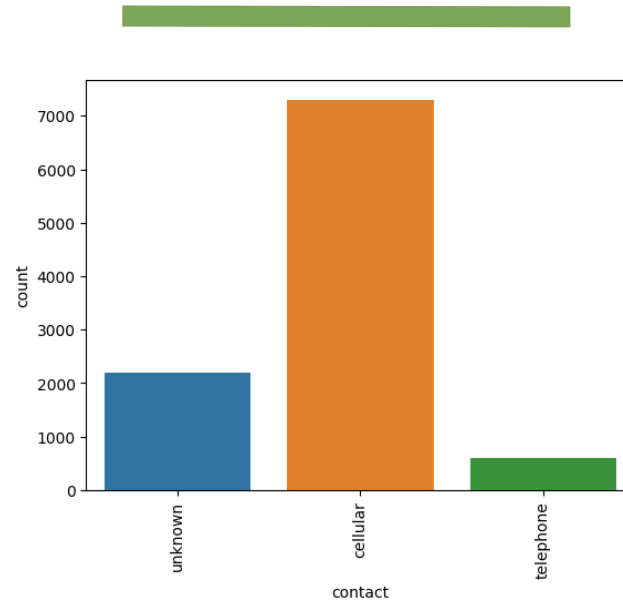
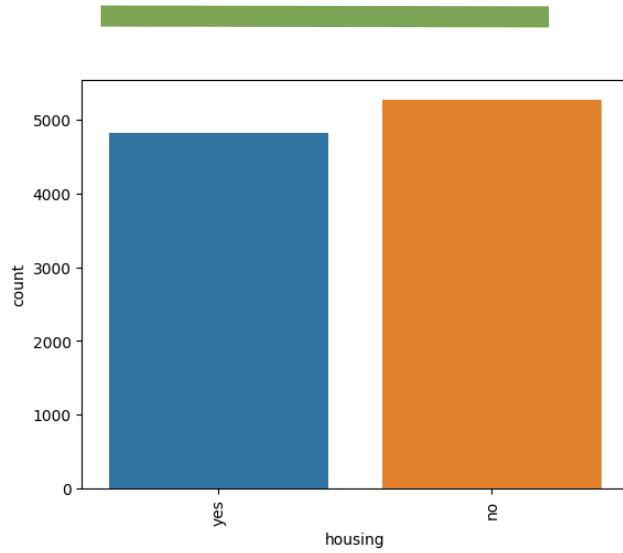
# اکتشافات داده‌ای



# اکتشافات داده‌ای



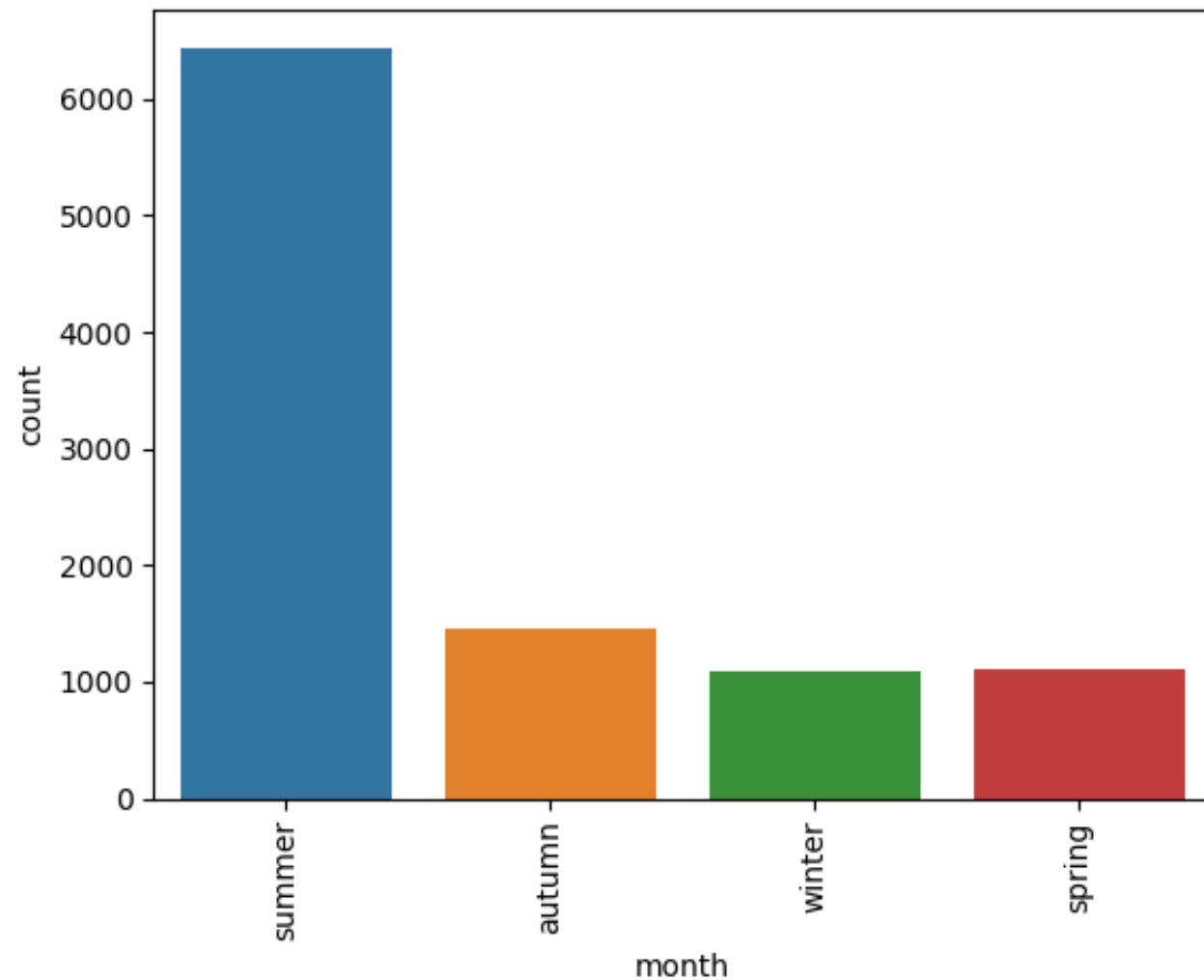
# اکتشافات داده‌ای





# اکتشافات داده‌ای

کوچک کردن تعداد بازه‌ها به فصل



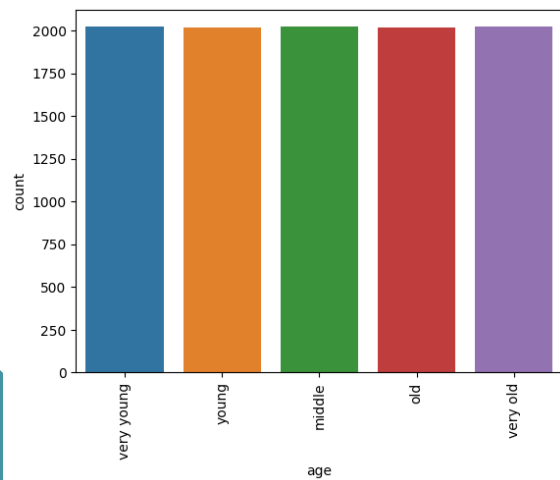
# اکتشافات داده‌ای - سبب‌بندی

Cut  
Qcut  
Natural Break

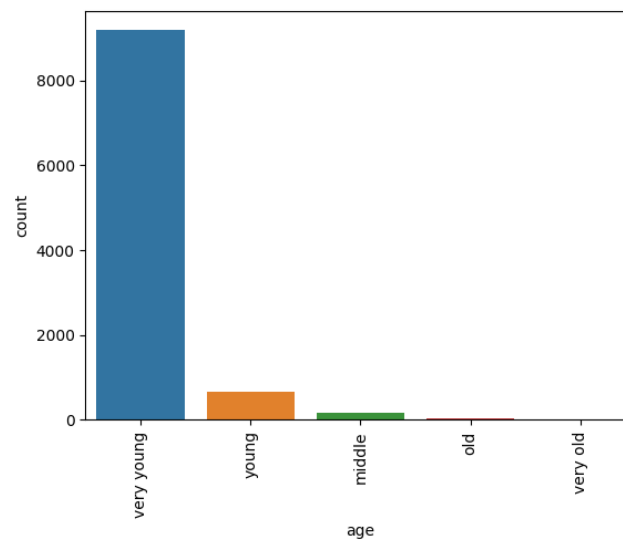


# سبب بندی

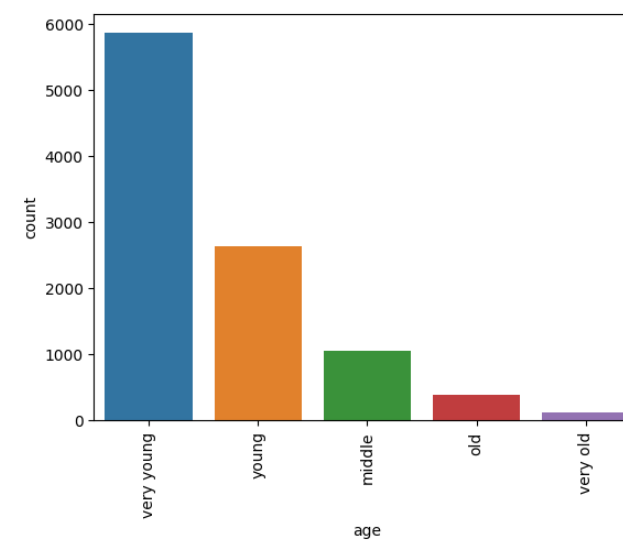
Qcut



CUT

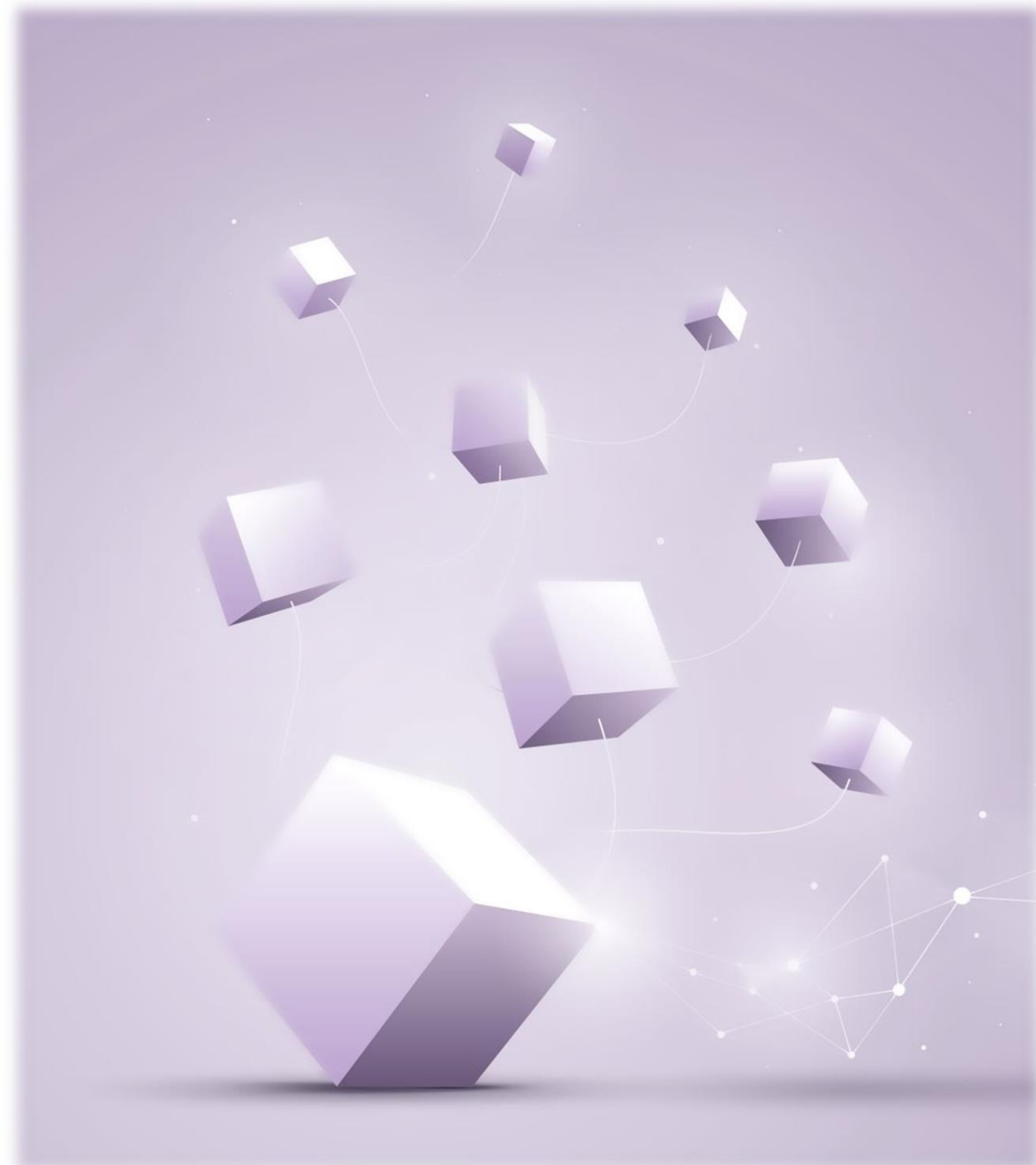


break



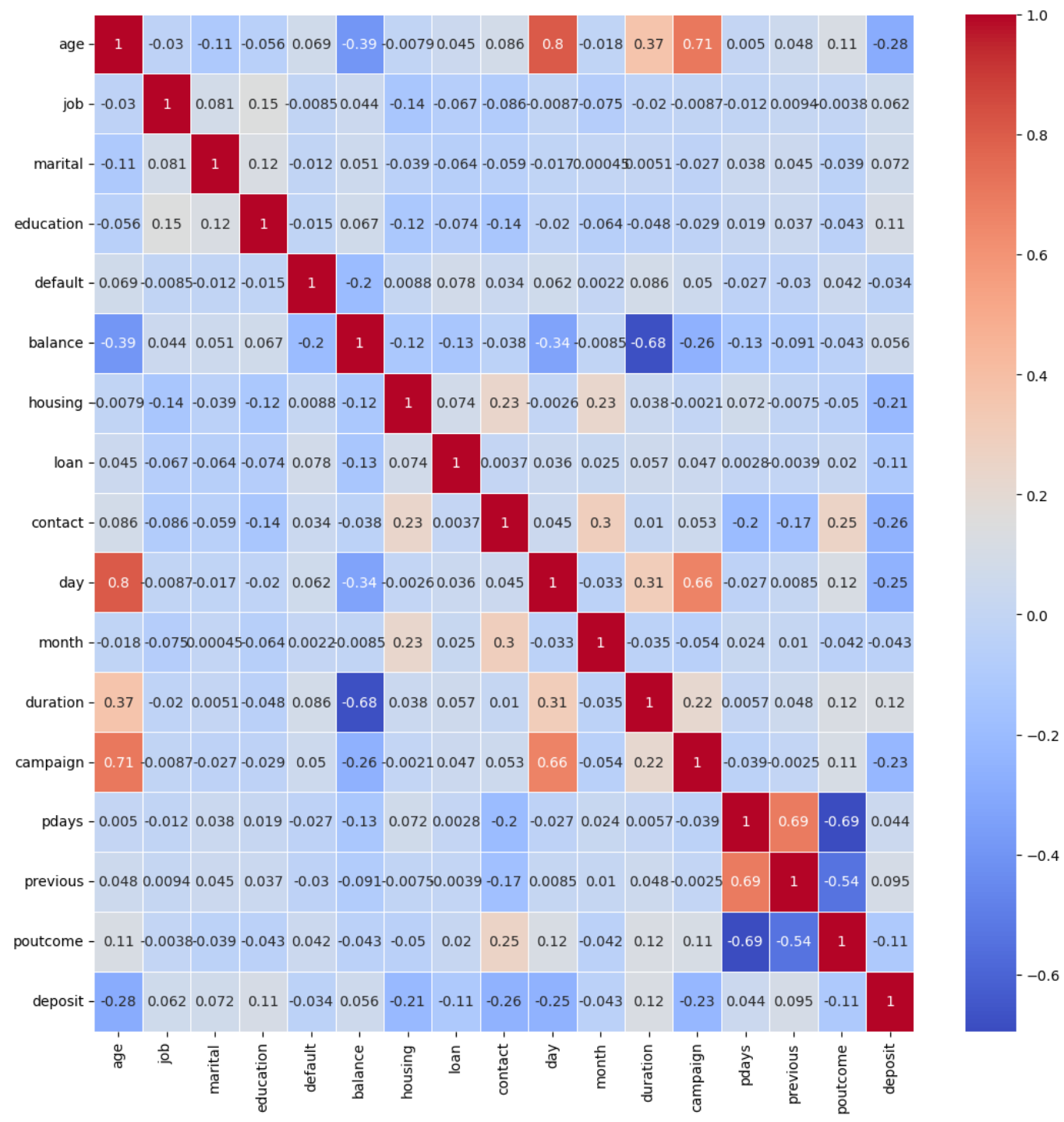
# اکتشافات داده‌ای - همبستگی

---



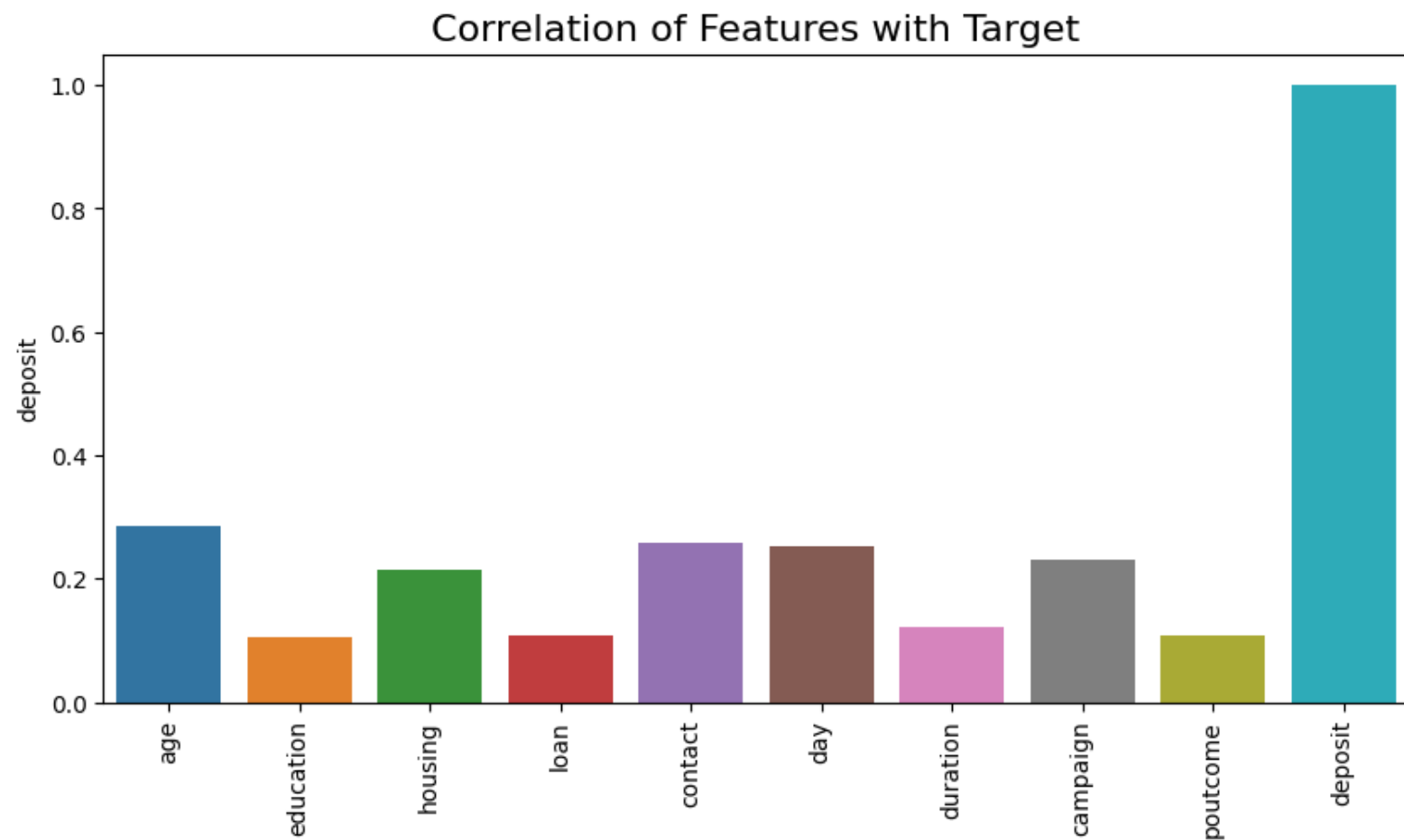
# اکتشافات داده‌ای - همبستگی

همبستگی ویژگی‌ها با یکدیگر



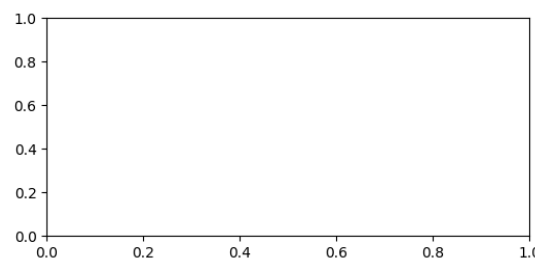
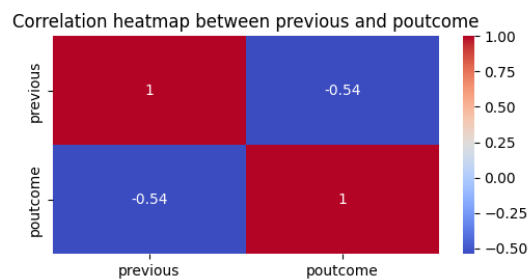
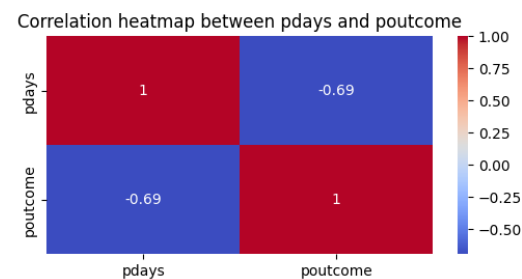
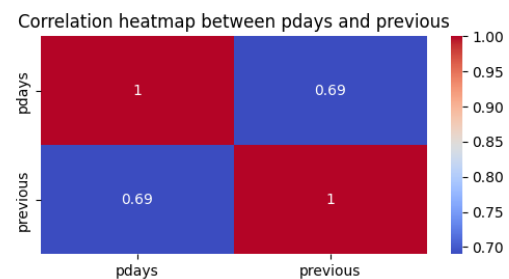
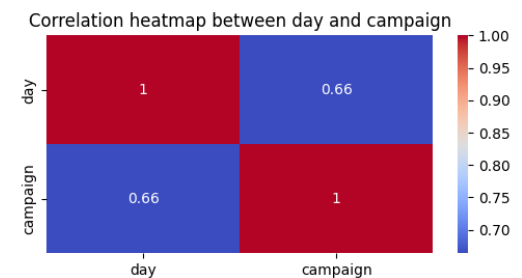
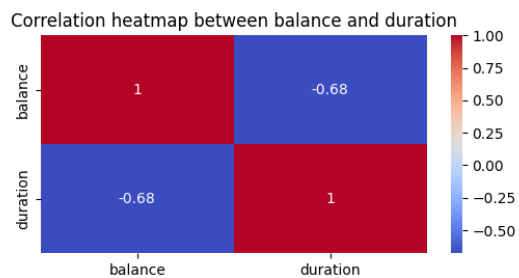
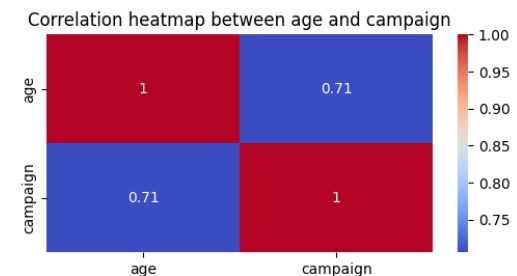
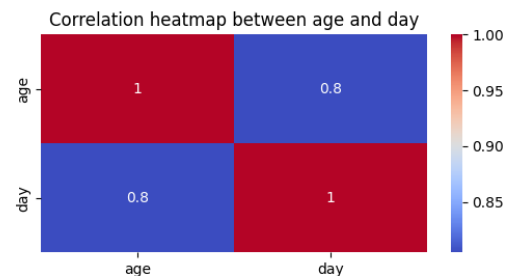
# اکتشافات داده‌ای - همبستگی

همبستگی ویژگی‌ها با متغیر هدف



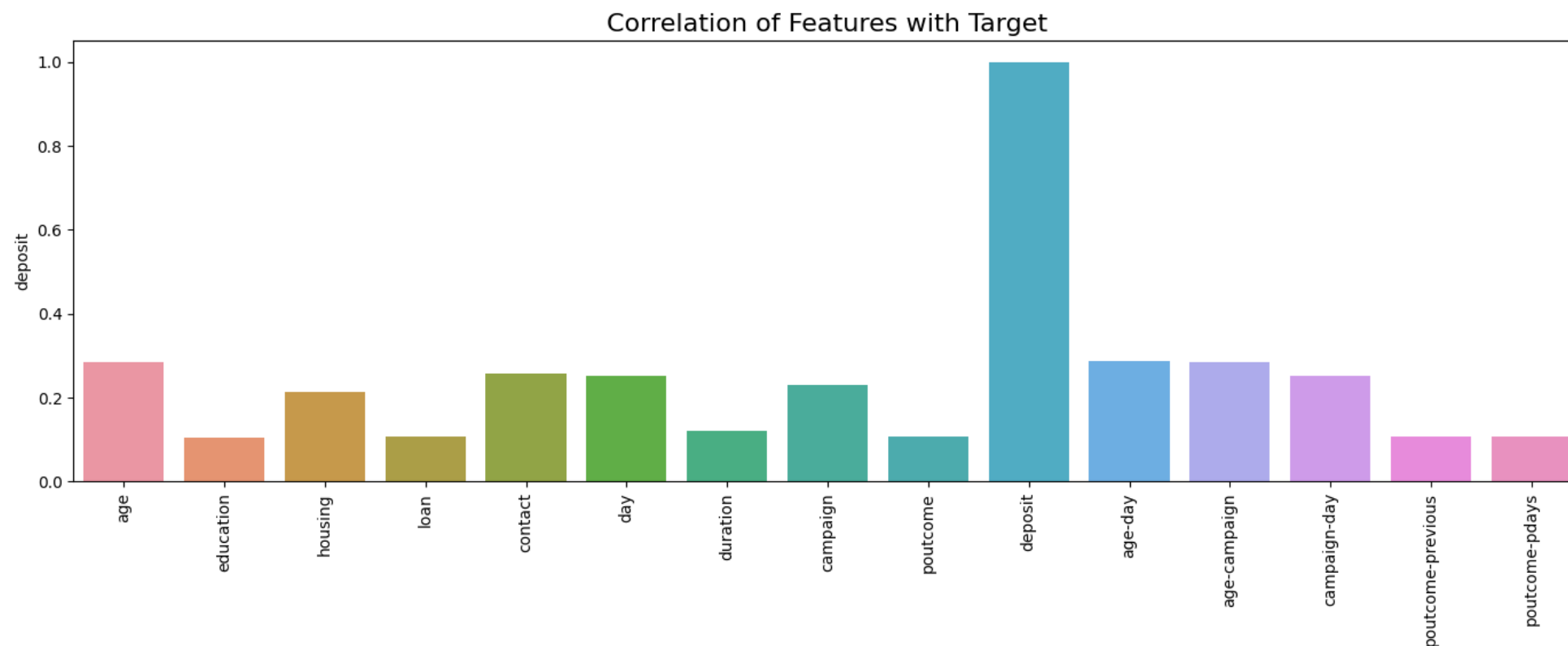
# اکتشافات داده‌ای - همبستگی

متغیر هایی با بیشترین همبستگی



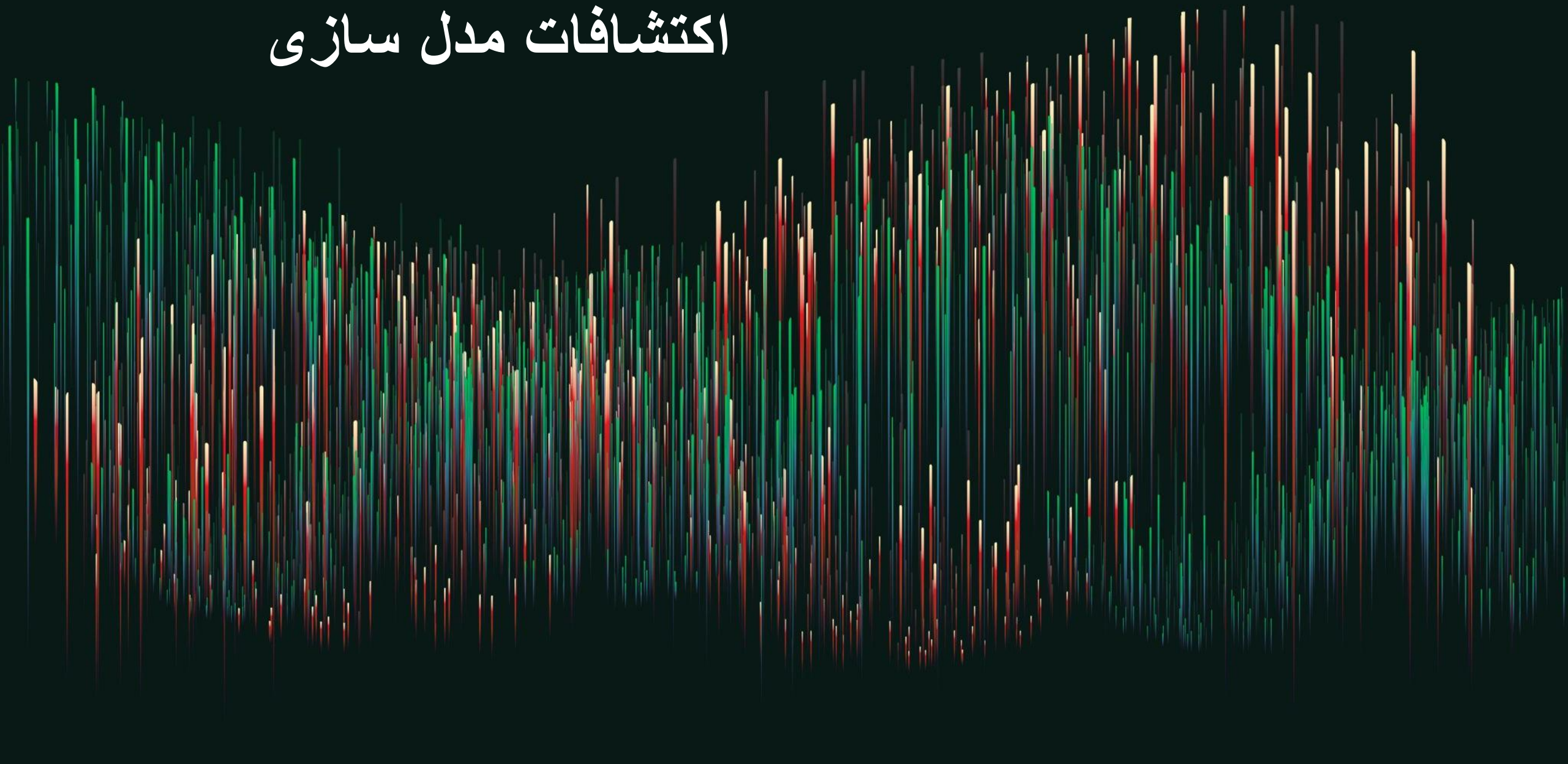
# اکتشافات داده‌ای - همبستگی

همبستگی ویژگی‌ها با متغیر هدف





# اکتشافات مدل سازی



## اکتشافات مدل سازی - فاز پیش مدل

تقسیم داده ها به  
۳۰ درصد تست و  
۷۰ درصد آموزش

cross  
validation

بالانس داده ها با  
standard  
scaler

## اکتشافات مدل سازی - انتخاب مدل

خوشه بندی

دسته بندی

## دسته بندی - اکتشافات مدل سازی

XGBoost

Neural Network

Random Forest

SVM

Stochastic  
Gradient  
Descent

Light Gradient  
Boosting  
Machine  
Classifier

## دسته بندی - اکتشافات مدل سازی

```
· test accuracy: 0.8222772277227722
      precision    recall  f1-score   support

         0         0.86    0.81    0.83     1109
         1         0.78    0.84    0.81      911

 accuracy          0.82     2020
  macro avg         0.82    0.82    0.82     2020
 weighted avg         0.82    0.82    0.82     2020
```



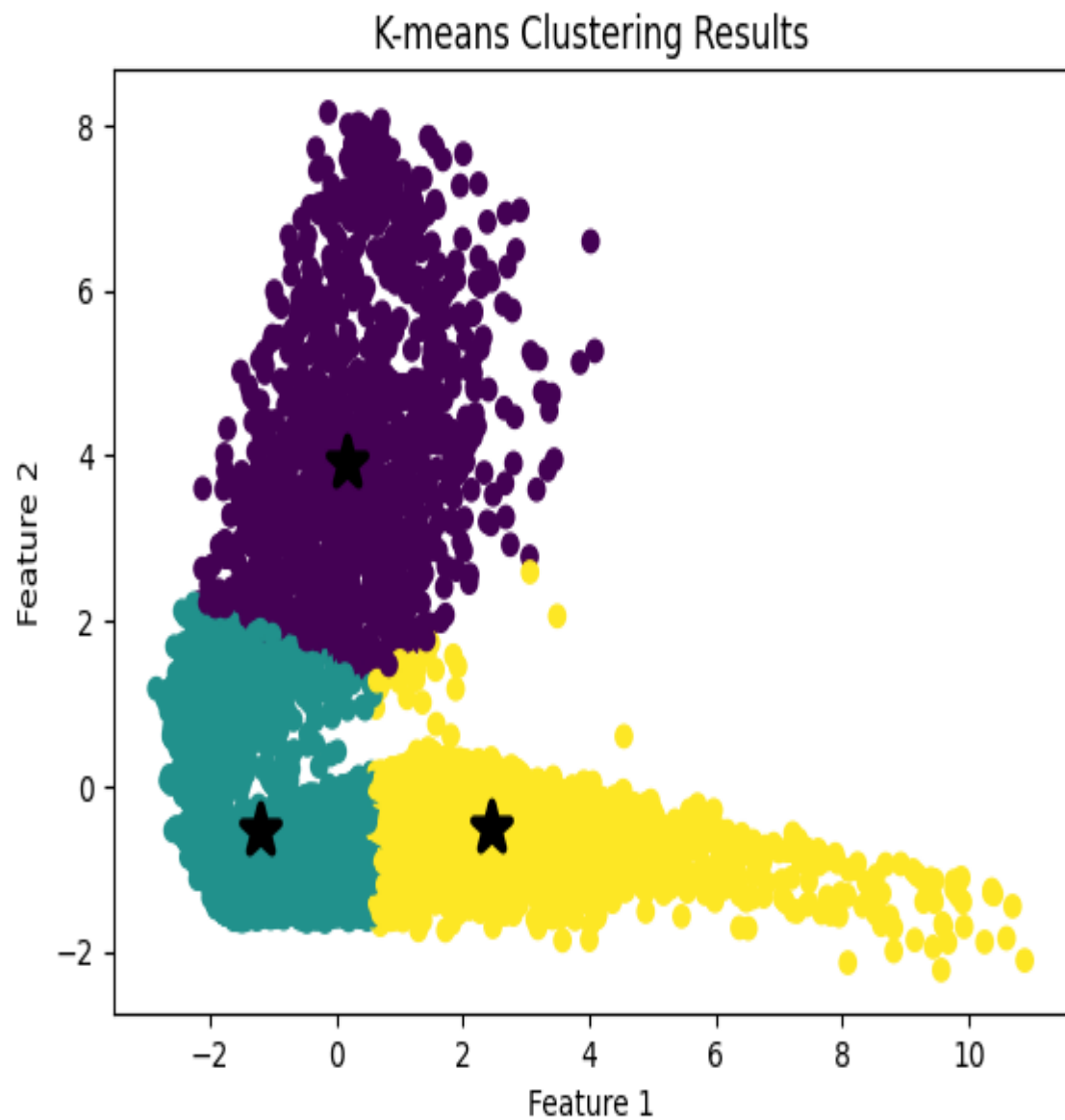
## اکتشافات مدل سازی-خوشه بندی مشتریان

---

با استفاده از K-means++

کاهش ویژگی ها به دوتا

روش PCA



## اکتشافات مدل سازی-خوشه بندی مشتریان

با استفاده از K-means++

کاهش ویژگی ها به دوتا

روش PCA





نتیج





# پیشنهادهای بیزنسی

## ماه برگزاری کمپین

بهترین ماه : ماه می

بیشترین رد تماس: ماه می

بهترین ماهها برای برگزاری کمپین به ترتیب: مارچ، سپتامبر، اکتبر

## ایجاد پرسشنامه

افزایش طول مکالمه با استفاده از پرسشنامه برای کاربران

مهمترین ویژگی استخراج شده در ریشه درخت

# مراجع

- [\[Moro et al., 2014\] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014](#)
- [S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014](#)
- [S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. \(Eds.\), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS. \[bank.zip\]](#)
- [archive.ics.uci.edu/ml/datasets/Bank+Marketing](http://archive.ics.uci.edu/ml/datasets/Bank+Marketing)
- <https://www.kaggle.com/datasets/janiobachmann/bank-marketing-dataset/code?datasetId=4471>

