

به نام خداوند دانش و پویایی

داده کاو: امیرحسین کیانی

موضوع: تحلیل فروش یک شرکت دانش بنیان

Data miner: AmirHossein Kiyani

GITHUB: <https://github.com/ahkiyani>

LINKEDIN: <https://www.linkedin.com/in/amirhossein-kiyani1381>

G-MAIL: <https://amirho3einkiyani.2002@gmail.com>

Topic: Analysis of a Drug

مراحل کار قدم به قدم:

۱. دانلود فایل داده بوسیله برنامه داتلودر

۲. خواندن و ذخیره فایل بصورت دیتافریم برای انجام پیش پردازش...

۳. نگاه به شمای کلی داده: داده درمورد ویژگی آدمها و بیماران و میزان کلسترول،

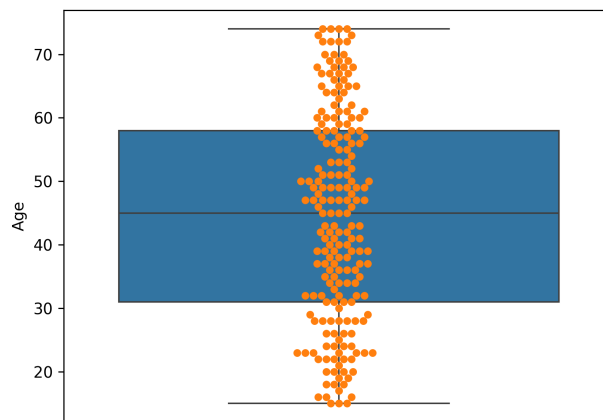
فشار خون و سدیم پتاسیم و داروی پیشنهادی هست

۴. باتوجه به داده ها نیاز به درخت تصمیم هستش و اینکه تشخیص بدیم چه دارویی برای چه کسی مناسبه؟

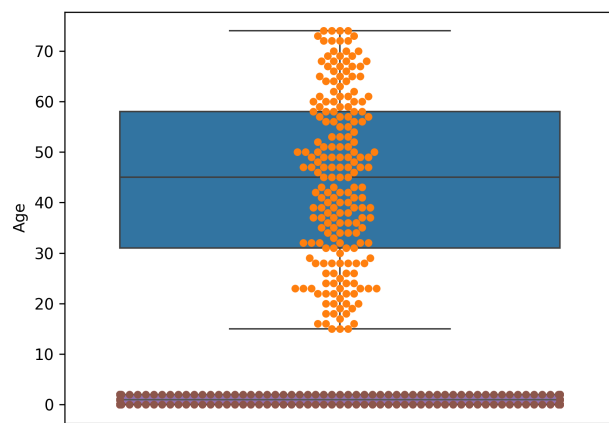
۵. پیش پردازش شامل حذف داده های تکراری، تصحیح داده های خطا و پیش بینی داده های گم شده

۶. نامگذاری ویژگی ها و نرمالسازی آنان

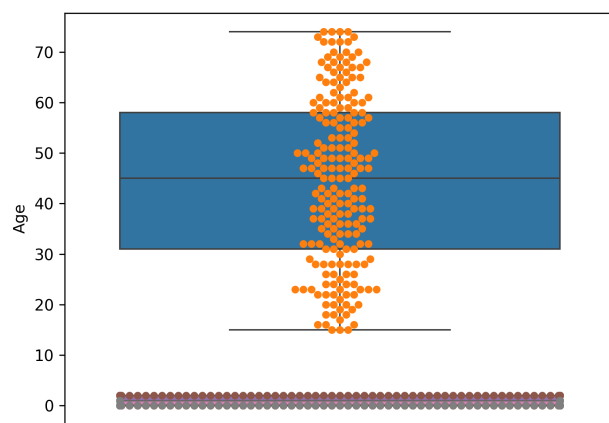
۷. تحلیل داده خام و نمودارهای آن و بررسی نویزها



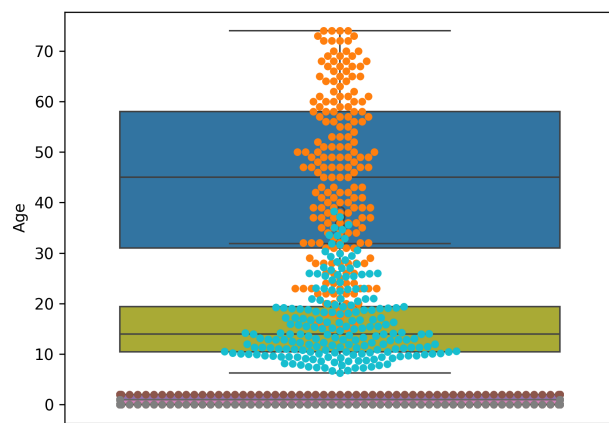
سن خام



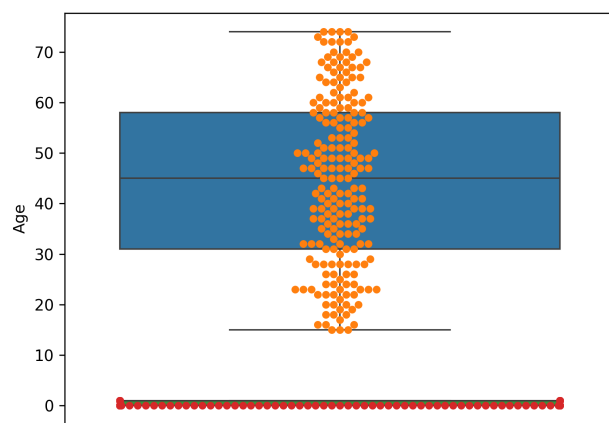
فشارخون خام



كلسترول خام



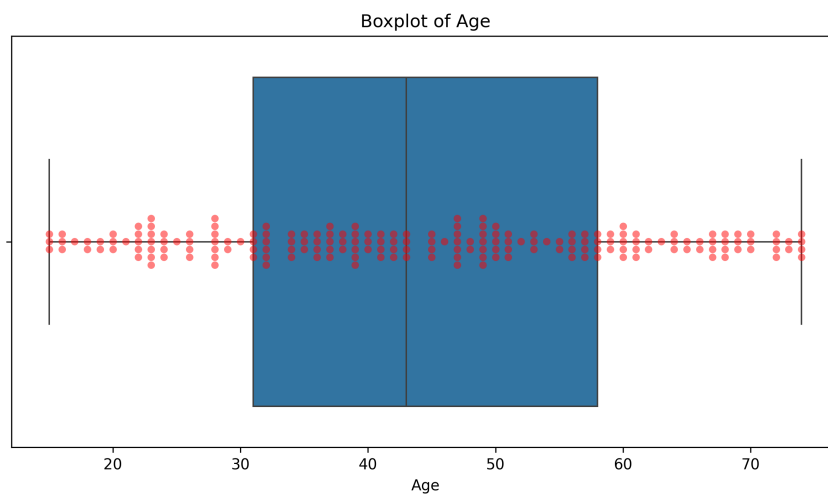
سدیم پتاسیم خام



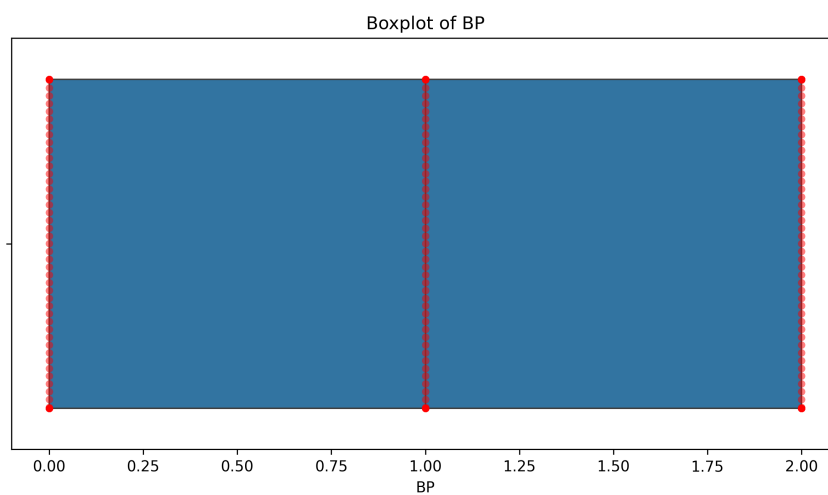
جنسیت خام

فقط فیچر سدیم پتاسیم دارای نویز است!

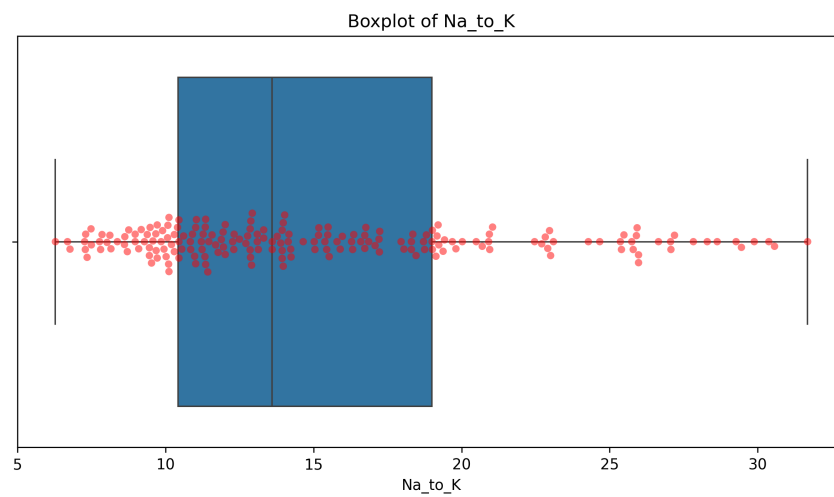
۸. پاکسازی داده و حذف نویزها، تکراری ها و در صورت نیاز حذف خالی ها یا پر کردن آنان با مقادیر منطقی



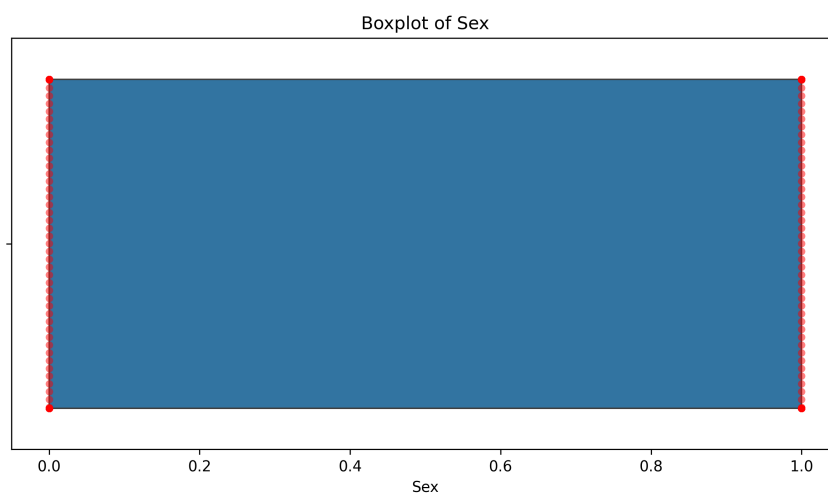
سن پاکسازی شده



فشارخون پاکسازی شده



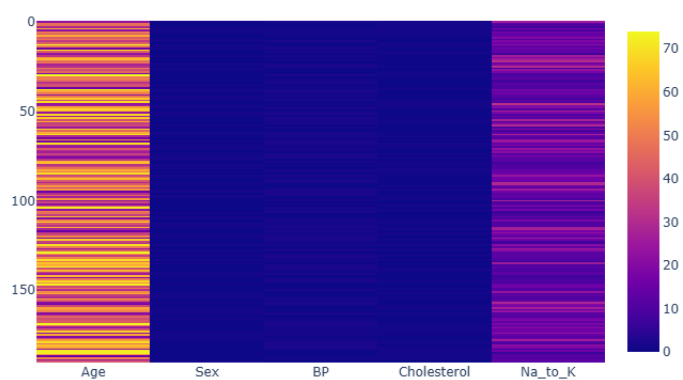
سدیم پتاسیم پاکسازی شده



جنسیت پاکسازی شده

۹. بررسی هم بستگی و هیت مپ هم بستگی

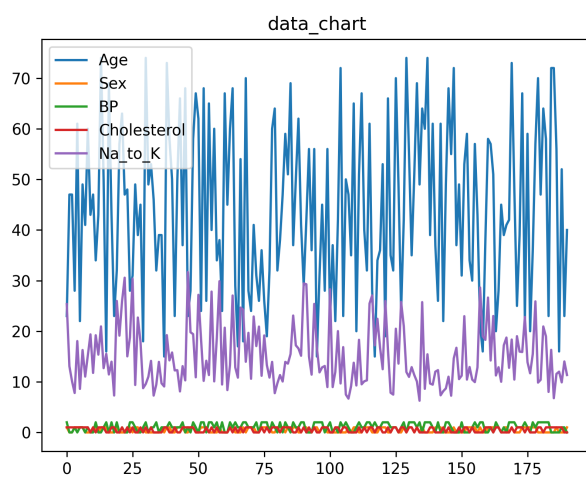
HeatMap Correlation



هیت مپ هم بستگی فیچرها

باتوجه به هم بستگی نمیتوان نتیجه دقیقی گرفت چون هم بستگی مفیدی وجود ندارد

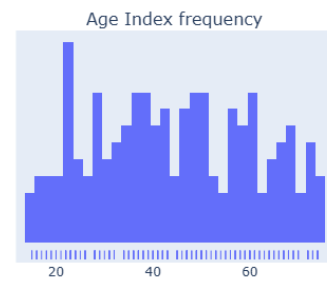
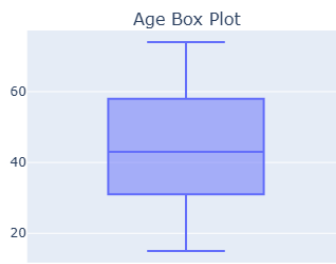
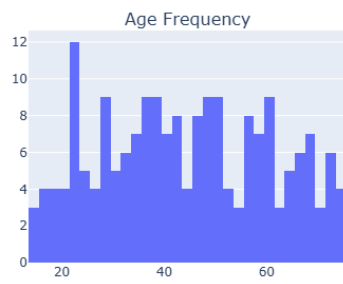
۱۰. رسم نمودار کلی از داده



نمودار کلی داده

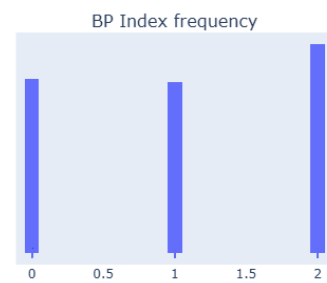
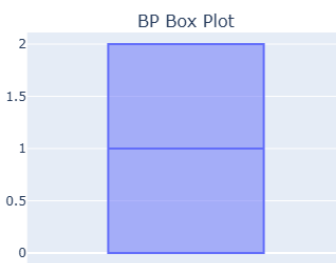
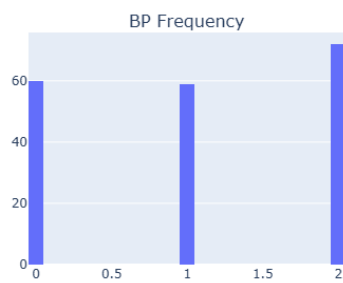
۱۱. رسم نمودار سه تایی برای تک تک فیچرها

Three charts for feature: Age



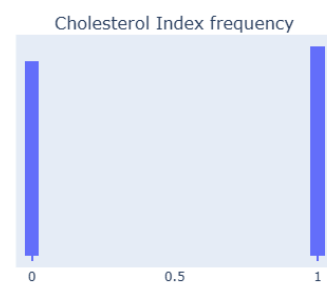
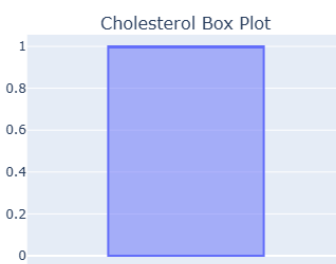
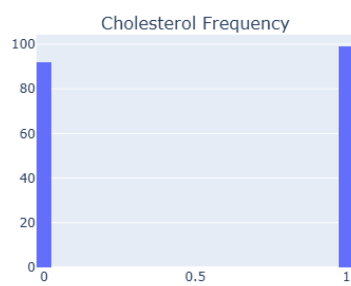
سن

Three charts for feature: BP



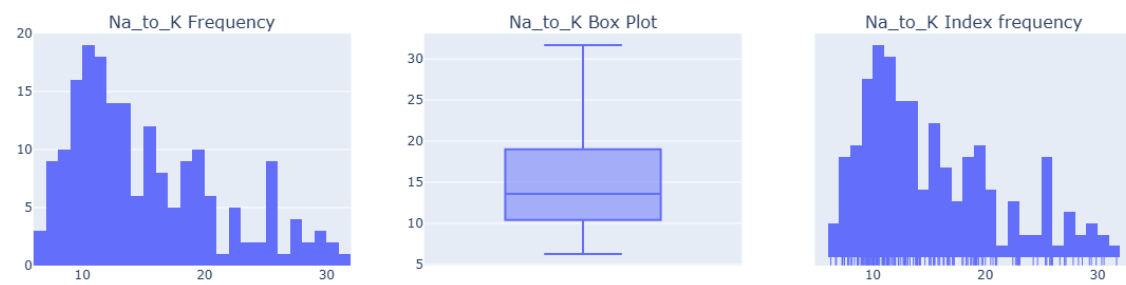
فشار خون

Three charts for feature: Cholesterol



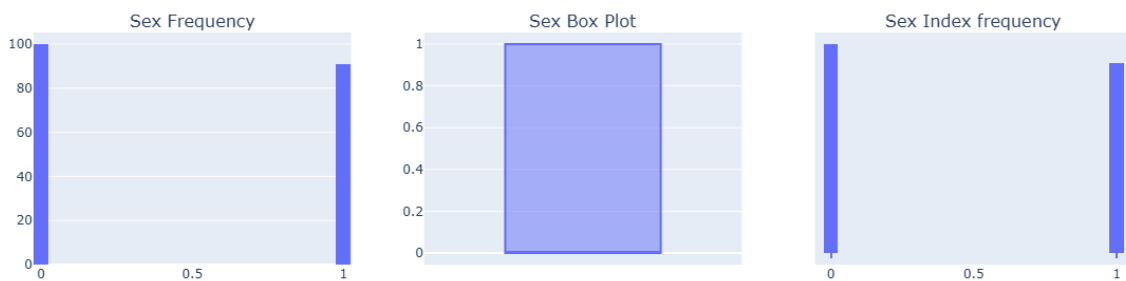
کلسترول

Three charts for feature: Na_to_K



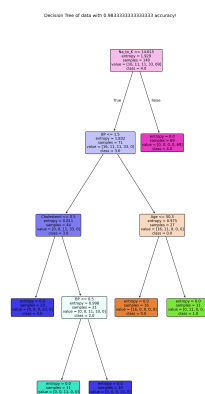
سدیم پتاسیم

Three charts for feature: Sex



جنسیت

با استفاده از درخت تصمیم و مدل سازی با روش آنتروپی سعی میکنیم تشخیص دهیم در چه مواردی کدام داروها باید تجویز شوند:



درخت تصمیم با دقت تقریباً ۹۸ درصدی

داروی نوع پنجم برای همه بیماران علی الخصوص کسانی که سدیم پتاسیم بیشتر از ۱۴.۶ دارند!

داروی نوع چهارم برای تمام بیماران با سدیم پتاسیم کمتر مساوی ۱۴.۶

و کسانی که فشارخون کمتر از ۱.۵ و کلسترول کمتر از ۰.۵

و کسانی که فشارخون بیشتر از ۰.۵ و کلسترول بیشتر از ۰.۵ دارند!

داروی نوع سوم برای تمام کسانی که سدیم پتاسیم زیر ۱۴.۶، فشارخون کمتر از ۱.۵ و کلسترول بیشتر از ۰.۵ دارند!

داروی نوع اول برای کسانی که سدیم پتاسیم کمتر از ۱۴.۶ و سن کمتر از ۵۰ دارند!

داروی نوع دوم برای کسانی که سدیم پتاسیم کمتر از ۱۴.۶ و سن بیشتر از ۵۰ دارند!

تمام!

Github: <https://github.com/ahkiyani>

Linkedin: <https://www.linkedin.com/in/amirhossein-kiyani1381>

G-mail: <https://amirho3einkiyani.2002@gmail.com>