به نام خدا

پروژهٔ درس آمار و احتمال مهندسی

عنوان پروژه:

تحلیل توصیفی دهک‌های درآمدی و رفتارهای مالی ایرانیان (۱۴۰۲–۱۳۹۸)

Project Title:

Descriptive Analysis of Income-Decile Classification and Financial Behaviors among Iranians (2019–2023)

پژوهشگر: **امیررضا شیرکوند**

Researcher: **Amirreza Shirkavand**

استاد راهنما: **دکتر سعید غضنفری‌راد**

Course Instructor: **Prof. Saeed Ghazanfari-Rad**

Amirreza Shirkavand

فهرست مطالب

۱. چکیده ۳

۲. معرفی ۳

۳. مجموعه‌دادهٔ مورد بررسی ۴

برخورد با مقادیر مفقود ۶

۴. تحلیل داده با آمار توصیفی ۷

۱.۴. بررسی جمعیت‌شناختی ۸

۲.۴. توزیع دهک‌های درآمدی و رفتارهای مالی ۱۰

۳.۴. اشتغال و اموال افراد در هر دهک ۱۲

۵. بحث و توضیح نتایج ۱۴

میانگین مبلغ ماهانهٔ تراکنش‌های بانکی در دهک‌های مختلف چه تفاوتی با یکدیگر دارند؟ ۱۴

آیا در طول سال‌های ۱۳۹۸ تا ۱۴۰۲، یک روند صعودی در میانگین تراکنش‌های بانکی مشاهده می‌شود؟ ۱۴

کدام‌یک از متغیرها بیشترین ارتباط را با دهک درآمدی اشخاص دارند؟ ۱۴

۶. نتیجه‌گیری ۱۴

۷. بیانیهٔ استفاده از ابزار هوش مصنوعی ۱۵

۸. منابع ۱۵

۹. واژگان ۱۵

۱. چکیده

پروژهٔ حاضر با استفاده از روش‌های آمار توصیفی[1] به تحلیل مجموعه‌دادهٔ[2] رفاه ایرانیان با حدود ۱.۷ میلیون رکورد می‌پردازد که برای هدفمند کردن پرداخت یارانه‌های دولت به‌کار می‌رود. در این مطالعه بررسی می‌کنیم که رفتارهای مالی (تراکنش‌های ماهانهٔ بانکی)، دارایی‌ها و اموال (موجودی حساب بانکی، وسایل نقلیه و پرتفوی بورسی) و ویژگی‌های جمعیت‌شناختی[3] (از جمله سن، جنسیت و محل سکونت شهری یا روستایی) در میان دهک‌های درآمدی در طول سال‌های هدف، چه الگویی دارند. برای انجام محاسبات و مصورسازی آماری از کتاب‌خانه‌های pandas و matplotlib در زبان برنامه‌نویسی python استفاده شد. نتایج نشان داد که میان دهک درآمدی و میزان تراکنش‌های ماهانهٔ کارت بانکی رابطهٔ مثبتی وجود دارد و حجم تراکنش‌ها در طی سال‌های ۱۳۹۸ تا ۱۴۰۲ روندی افزایشی داشته است؛ اما ارزش واقعی آنها (با حذف تورم) ثابت مانده است. همچنین، متغیرهایی مانند شهرنشینی، ارزش سبد سهام، بیمه‌پرداز بودن و داشتن جواز کسب نسبت به سایر متغیرهای بررسی‌شده همبستگی مثبت بیشتری با دهک درآمدی نشان دادند، هرچند شدت این همبستگی‌ها ضعیف بود.

Abstract

This project applies descriptive statistical methods to analyze the Iranian Welfare Dataset, which contains approximately 1.7 million records and is used to support the government's means-testing system for targeted subsidy allocation. The study examines the patterns of financial behavior (monthly bank card transactions), assets (including bank account balances, vehicles, and stock portfolios), and demographic characteristics (including age, gender, and urban or rural residence) across different income deciles over the study period. Statistical analysis and data visualization were performed using the pandas and matplotlib libraries in the Python programming language. The results showed a positive relationship between income decile and the volume of monthly bank card transactions. Furthermore, transaction volumes exhibited an increasing trend from the years 1398 to 1402 (2019–2023); however, their real values, after adjusting for inflation, remained relatively constant. In addition, variables such as urban residence, stock portfolio value, insurance status, and business license ownership showed relatively stronger positive correlations with income deciles than the other variables examined, although the overall strengths of these correlations were weak.

۲. معرفی

هدفمندسازی یارانه‌ها در ایران به کمک «الگوریتم آزمون وسع»[4] و تقسیم جمعیت به ده دهک درآمدی (دهک ۱ کمترین و دهک ۱۰ بیشترین درآمد را دارد) انجام می‌شود (محمدی، ۱۴۰۵). متغیرهای متعددی در این الگوریتم تاثیر داده می‌شوند که در ادامه معرفی می‌شوند. بررسی تغییرات و روندهای موجود در داده‌های این مجموعه‌داده، می‌تواند نقش مهمی در تصمیمات و سیاست‌گذاری‌های اقتصادی ایفا کند.

[1] Descriptive Statistics
[2] Dataset
[3] Demographics
[4] Means Test Algorithm

در این پروژه با استفاده از شیوه‌های آمار توصیفی به سوالات زیر پاسخ می‌دهیم:

۱. میانگین مبلغ ماهانهٔ تراکنش‌های بانکی در دهک‌های مختلف چه تفاوتی با یکدیگر دارند؟

۲. آیا در طول سال‌های ۱۳۹۸ تا ۱۴۰۲، یک روند صعودی در میانگین تراکنش‌های بانکی مشاهده می‌شود؟

۳. کدام‌یک از متغیرها بیشترین ارتباط را با دهک درآمدی اشخاص دارند؟

فصل ۳ به معرفی مجموعه‌دادهٔ مورد استفاده و متغیرهای موجود در آن اختصاص داده شده و در فصل ۴ به تحلیل توصیفی داده‌های آن می‌پردازیم. برای تحلیل توصیفی از زبان برنامه‌نویسی Python و کتاب‌خانه‌های[5] مرتبط استفاده شده که در ابتدای همان فصل به آنها اشاره شده است. در فصل ۵ به سوالاتی که قبل‌تر مطرح شد پاسخ می‌دهیم.

۳. مجموعه‌دادهٔ مورد بررسی

وزارت کار، تعاون و رفاه اجتماعی ایران داده‌های مرتبط به خانوارهای کشور از منابع گوناگون گردآوری و در یک پایگاه داده به نام پایگاه رفاه ایرانیان ذخیره می‌کند و از آن در خدماتی مانند استحقاق‌سنجی در خدمات حمایتی و تحلیل‌های کلان استفاده می‌کند. این پایگاه داده از تجمیع ۵۰ منبع داده‌ای تشکیل شده است. در فاز نخست، ۲۵ منبع با بیش از ۶۰ جدول داده به طور کامل تجمیع شده و ساختار اصلی پایگاه اطلاعات رفاه ایرانیان بر اساس آن شکل گرفته است. بیش از ۳ میلیارد رکورد داده در پایگاه‌دادهٔ نهایی ذخیره و تجمیع شده است.

برخی از این سازمان‌ها عبارتند از: سازمان تامین اجتماعی، سازمان بیمه سلامت، صندوق بازنشستگی کشوری، صندوق بیمهٔ اجتماعی کشاورزان، روستاییان و عشایر، سازمان بهزیستی کشور، کمیتهٔ امداد امام خمینی، سازمان ثبت احوال، اتحادیه املاک، سازمان آموزش فنی‌وحرفه‌ای، وزارت آموزش‌وپرورش، وزارت علوم، تحقیقات و فناوری، نیروی‌انتظامی، بانک مرکزی، سازمان هدفمندسازی یارانه‌ها، سازمان امورمالیاتی، وزارت بهداشت، بنیاد شهید و امور ایثارگران و شرکت پست ایران. این وزارتخانه نمونهٔ ۲ درصدی این داده‌ها را پس از حذف اطلاعات محرمانه، برای استفادهٔ پژوهشگران به صورت عمومی منتشر می‌کند (معرفی داده‌های بانک رفاه ایرانیان، ۱۴۰۳). مجموعه‌داده به‌صورت یک فایل به فرمت csv[6] با کدگذاری UTF-8 در دو قسمت[7] فایل فشردهٔ rar[8] ارائه شده است. این فایل شامل ۱,۶۹۷,۸۱۶ سطر داده بوده و هر سطر نمایان‌گر یک فرد در سامانهٔ رفاه و آزمون وسع است. مجموعه‌دادهٔ منتشر شده دارای ۴۸ ستون بوده که جدول ۱ به تشریح آنها پرداخته است.

آدرس مجموعه‌داده در وب‌سایت پایگاه رفاه ایرانیان:

https://refahdb.mcls.gov.ir/fa/downloaddata-دانلودداده

[5] Libraries
[6] Comma-Separated Values
[7] Part
[8] Roshan Archive

جدول ۱. شرح متغیرهای مجموعه‌داده

–	متغیر	توضیحات	نوع متغیر	واحد
1	id	شناسه	شناسهٔ عددی	—
2	Parent_Id	شناسهٔ سرپرست خانوار	شناسهٔ عددی	—
3	GenderId	جنسیت (مرد=۱، زن=۲)	کیفی اسمی	—
4	Age	سن	کمی گسسته	سال
5	ISBimarKhas	آیا بیماری خاص دارد؟	دودویی	—
6	IsMalool	آیا معلولیت دارد؟	دودویی	—
7	Malool_shedat	شدت معلولیت	کیفی ترتیبی	—
8	Dashboard_postalcode7Digits	هفت رقم ابتدای کد پستی (محل سکونت)	کیفی اسمی	—
9	isurban	شهری=۱ یا روستایی=۰	دودویی	—
10	SabteAhval_provincename	شهرستان محل سکونت	کیفی اسمی	—
11	SabteAhval_countyname	استان محل سکونت	کیفی اسمی	—
12	Has_SoeTaghzie	آیا سوء تغذیه دارد؟	دودویی	—
13	IsBehzisti_AfzayeshMostamari	تحت پوشش بهزیستی است؟	دودویی	—
14	IsKomite_AfzayeshMostamari	تحت پوشش کمیتهٔ امداد است؟	دودویی	—
15	IsKomite_AfzayeshMostamariSayer	آیا تحت پوشش کمیتهٔ امداد است؟	دودویی	—
16	TripCountAirPilgrimage_95to99	سفر زیارتی هوایی از سال ۹۵ تا ۹۹	کمی پیوسته	تعداد سفر
17	TripCountAirNonPilgrimage_95to99	سفر غیرزیارتی هوایی از سال ۹۵ تا ۹۹	کمی پیوسته	تعداد سفر
18	TripCountNonAirPilgrimage_95to99	سفر زیارتی غیرهوایی از سال ۹۵ تا ۹۹	کمی پیوسته	تعداد سفر
19	TripCountNonAirNonPilgrimage_95to99	سفر غیرزیارتی غیرهوایی از سال ۹۵ تا ۹۹	کمی پیوسته	تعداد سفر
20	Has_Saham_Edalat	آیا سهام عدالت دارد؟	دودویی	—
21	Decile	دهک درآمدی	کیفی ترتیبی	۱-۱۰
22	Percentile	صدک رفاهی	کیفی ترتیبی	۱-۱۰۰
23	HasMojavezSenfi	آیا مجوز صنفی دارد؟	دودویی	—
24	ISKarmanddolat_1402	آیا در سال ۱۴۰۲ کارمند دولت است؟	دودویی	—
25	IsRetired_Asli	آیا بازنشستهٔ اصلی است؟	دودویی	—
26	IsRetired_Tabaie	آیا بازنشستهٔ تبعی است؟	دودویی	—
27	is_bime_darman	آیا دارای بیمهٔ درمان است؟	دودویی	—
28	IsBimePardaz	آیا بیمه‌پرداز است؟	دودویی	—
29	MandehAval_1399	موجودی حساب در ابتدای سال ۱۳۹۹	کمی پیوسته	ریال

واحد	نوع متغیر	توضیحات	متغیر	-
ریال	کمی پیوسته	موجودی حساب در انتهای سال ۱۳۹۹	MandehAkhar_1399	30
ریال	کمی پیوسته	میانگین کارت‌به‌کارت ماهانه در ۱۳۹۹	CardPerMonth_1399	31
ریال	کمی پیوسته	مجموع واریزی در ۱۴۰۰	Variz_1400	32
ریال	کمی پیوسته	موجودی حساب در ابتدای سال ۱۴۰۰	MandehAval_1400	33
ریال	کمی پیوسته	موجودی حساب در انتهای سال ۱۴۰۰	MandehAkhar_1400	34
ریال	کمی پیوسته	مجموع خرید با کارت در ۱۴۰۰	CardPerMonth_1400	35
ریال	کمی پیوسته	مجموع خرید با کارت در ۱۳۹۸	CardPerMonth_1398	36
ریال	کمی پیوسته	مجموع خرید با کارت در ۱۴۰۱	CardPerMonth_1401	37
ریال	کمی پیوسته	مجموع خرید با کارت در ۱۴۰۲	CardPerMonth_1402	38
ریال	کمی پیوسته	میانگین کارت‌به‌کارت ماهانه در ۱۴۰۱	CardBeCardPerMonth_1401	39
ریال	کمی پیوسته	میانگین کارت‌به‌کارت ماهانه در ۱۴۰۲	CardBeCardPerMonth_1402	40
ریال	کمی پیوسته	گردش حساب ساتنا ماهیانه ۱۴۰۱	SatnaPerMonth_1401	41
ریال	کمی پیوسته	گردش حساب ساتنا ماهیانه ۱۴۰۲	SatnaPerMonth_1402	42
ریال	کمی پیوسته	گردش حساب پایا ماهیانه ۱۴۰۱	PayaPerMonth_1401	43
ریال	کمی پیوسته	گردش حساب پایا ماهیانه ۱۴۰۲	PayaPerMonth_1402	44
ریال	کمی پیوسته	ارزش خودروها	CarsPrice	45
تعداد	کمی گسسته	خودروها	CarsCount	46
ریال	کمی پیوسته	ارزش پورتفوی بورسی	Bourse_NetPortfoValue	47
ریال	کمی پیوسته	مجموع درآمدهای ثبتی	Daramad	48

برخورد با مقادیر مفقود[9]

بسیاری از متغیرها[10] مانند ISBimarKhas، IsMalool و IsRetired_Tabaie، دارای درصد بالایی از مقادیر مفقود هستند، زیرا این متغیرها تنها برای افرادی که آن وضعیت را دارند مقدار «بله» را ثبت می‌کنند و برای سایر افراد مقدار آنها خالی باقی می‌ماند؛ به بیان دیگر، مقدار مفقود به معنای «خیر» است.

[9] Missing Values
[10] Variables (Features/Characteristics)

پنج متغیر با بیشترین میزان مقادیر مفقود عبارت‌اند از:

- IsRetired_Tabaie با ٪۹۹.۵ دادهٔ مفقود
- IsKomite_AfzayeshMostamariSayer با ٪۹۹.۴ دادهٔ مفقود
- ISBimarKhas با ٪۹۹.۲ دادهٔ مفقود
- Malool_shedat با ٪۹۸.۴ دادهٔ مفقود
- IsMalool با ٪۹۸.۳ دادهٔ مفقود

در تحلیل‌های بعدی، این مقادیر مدیریت شده‌اند[11]؛ یعنی در متغیرهای دودویی، مقادیر مفقود با صفر (به معنای «خیر») جایگزین شده‌اند.

۴. تحلیل داده با آمار توصیفی

در این بخش، مجموعه‌داده را با استفاده از شیوه‌های آمار توصیفی بررسی می‌کنیم. برای این منظور، از کتاب‌خانه‌های pandas و matplotlib و برخی کتاب‌خانه‌های دیگر در زبان برنامه‌نویسی Python استفاده شده است و تمام نمودارها با استفاده از کدهای ارائه شده در پیوست، قابلیت بازتولید دارند. مشخصات سیستم‌عامل و محیط برنامه‌نویسی (IDE[12]) استفاده شده برای تحلیل داده و رسم نمودارها به شرح زیر است:

OS[13]: Windows 11 Pro 21H2 (Build: 22000.2538)

System Type: 64-bit operating system, x64-based processor

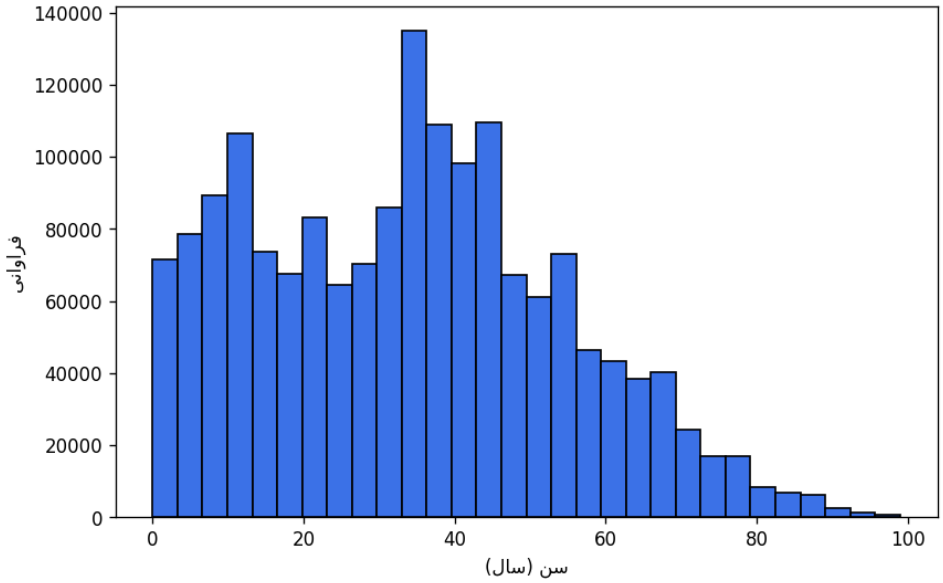
IDE: Visual Studio Code (Version: 1.108.2)

Programming Language: Python (Version: 3.13.1)

Used Libraries: pandas (2.3.0), matplotlib (3.10.3), scipy (1.16.0), arabic-reshaper (3.0.0), python-bidi (0.6.6)

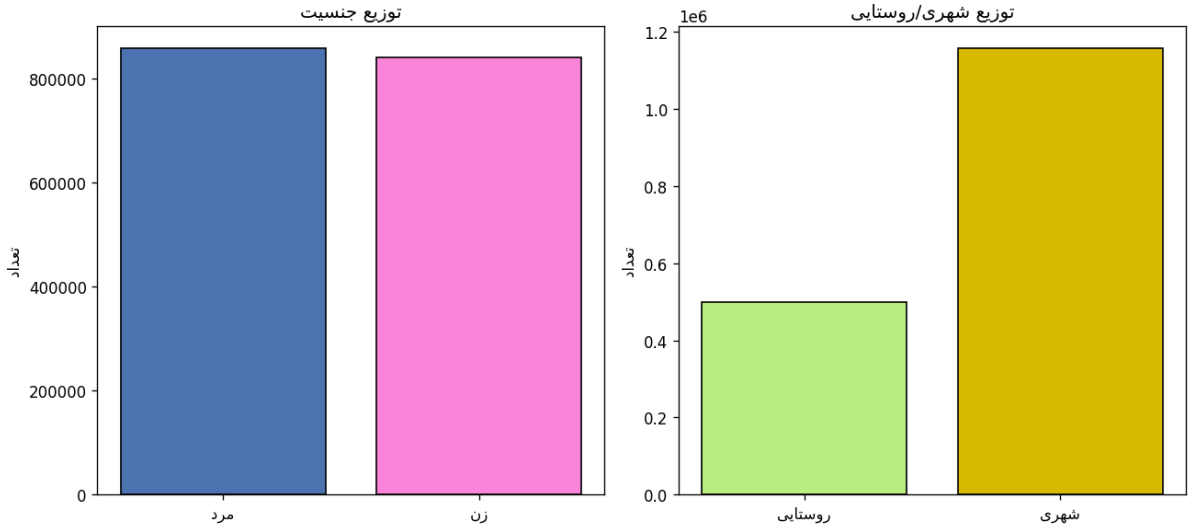
[11] Handling Missing Values
[12] Integrated Development Environment
[13] Operating System

۱.۴ . بررسی جمعیت‌شناختی



توزیع سن افراد[۱۴]نمودار ۱. هیستوگرام

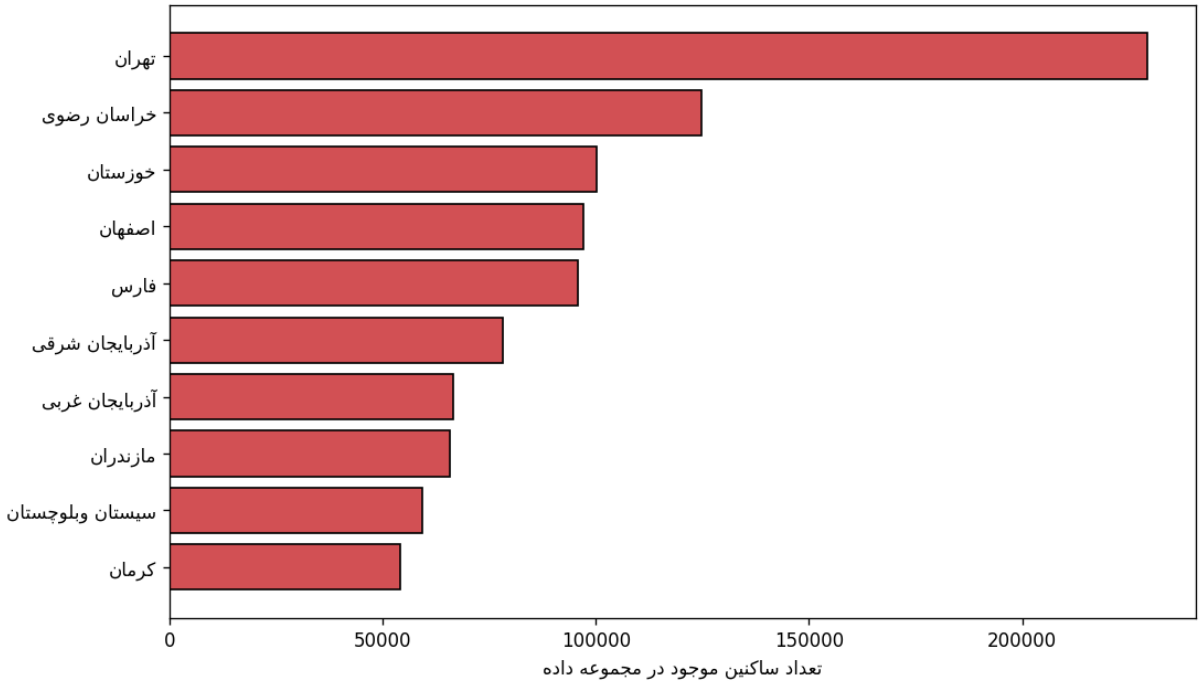
توزیع سنی افراد دارای چولگی به راست[۱۵] است. همچنین میانگین[۱۶] سن آنها حدود ۳۴.۱ و میانهٔ[۱۷] سن آنها ۳۴ سال با انحراف معیار[۱۸] ۲۰.۳ سال برآورد می‌شود.



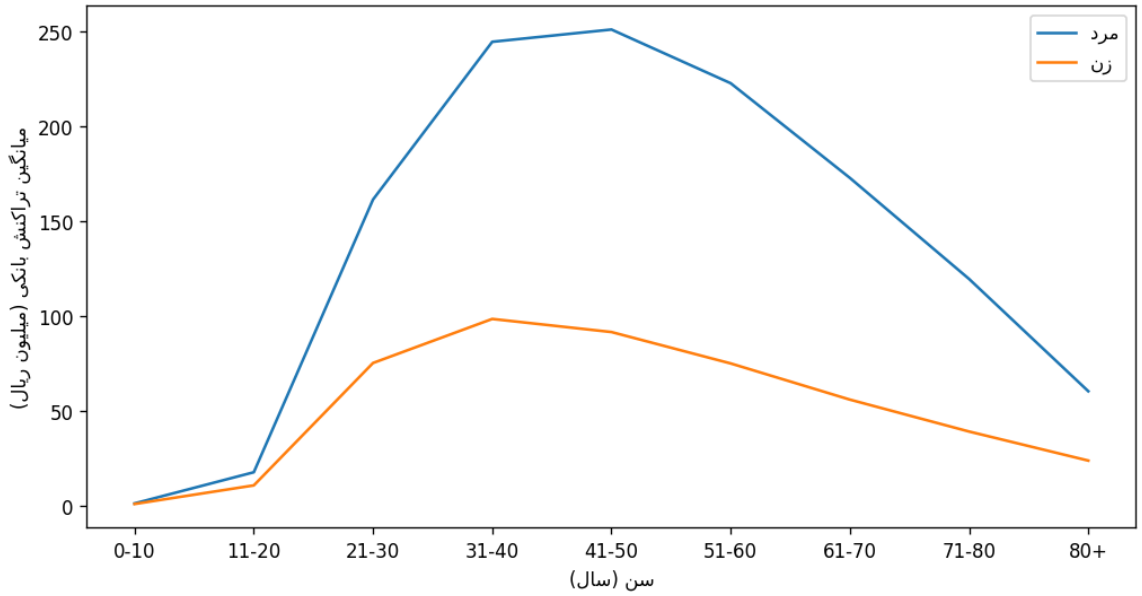
نمودار ۲. ساختار جمعیت‌شناختی (جنسیت و شهری/روستایی)

[14] Histogram
[15] Right-Skewed Distribution
[16] Mean
[17] Median
[18] Standard Deviation

تعداد مردان و زنان تقریباً برابر است (مردان اندکی بیشتر هستند). حدود ۷۰٪ افراد ساکن مناطق شهری هستند و سایر افراد در روستاها سکونت دارند.



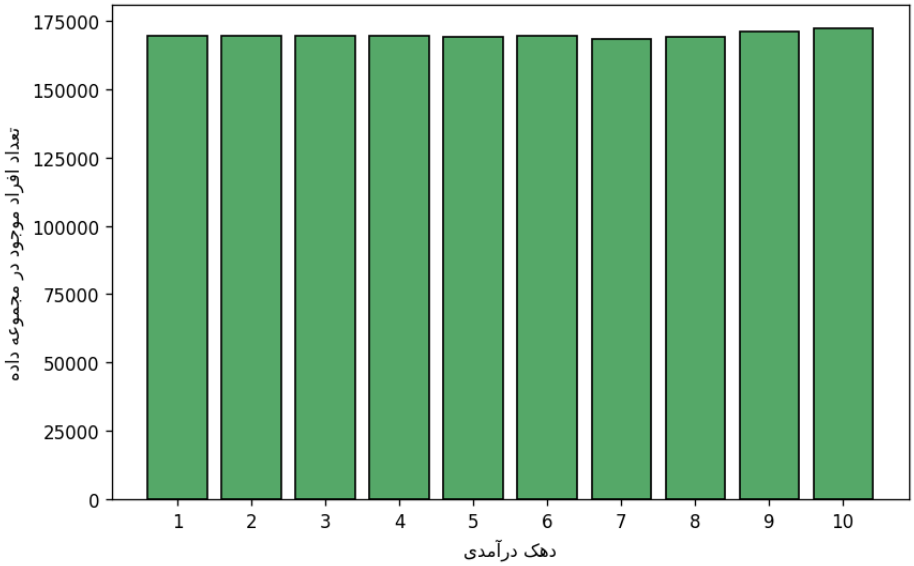
نمودار ۳. پرجمعیت‌ترین استان‌ها



نمودار ۴. میانگین ماهانهٔ تراکنش کارت بانکی در مردان و زنان به تفکیک سن

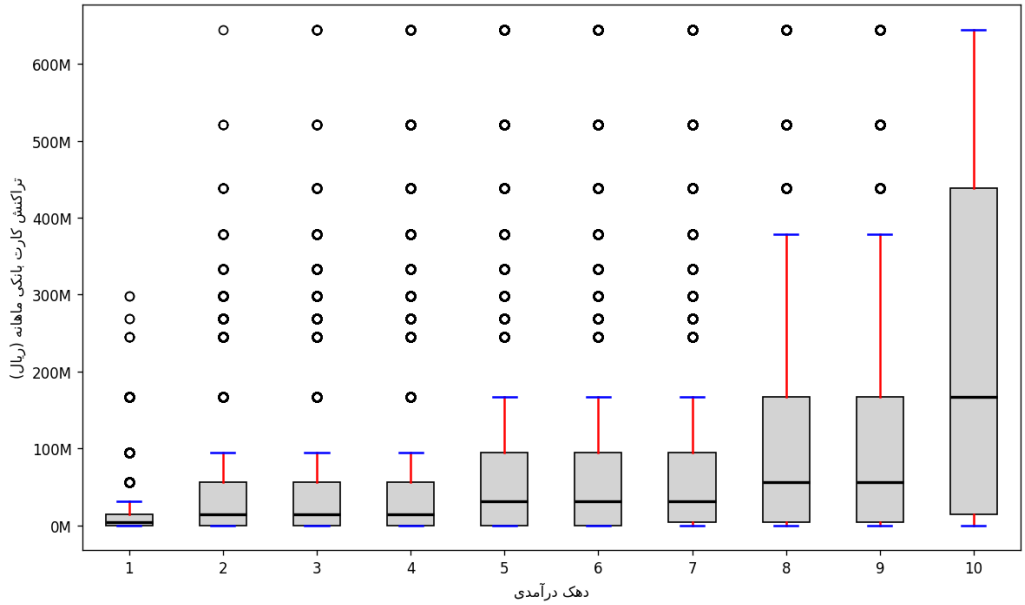
به‌نظر می‌رسد مردان اغلب به‌دلیل بر عهده داشتن سرپرستی خانوار و همچنین اشتغال بیشتر، در تمام دوران زندگی خود بیشتر از زنان تراکنش مالی دارند. همچنین نمودار ۴ نشان می‌دهد که احتمالاً مخارج - و به طبع، درآمد - افراد از سن ۲۰ سالگی شروع می‌شود که صعود آن در زنان تا ۴۰ سالگی و در مردان تا ۵۰ سالگی است؛ سپس یک روند اکیداً نزولی در تراکنش‌های هردو جنس دیده می‌شود.

۲.۴ . توزیع دهک‌های درآمدی و رفتارهای مالی



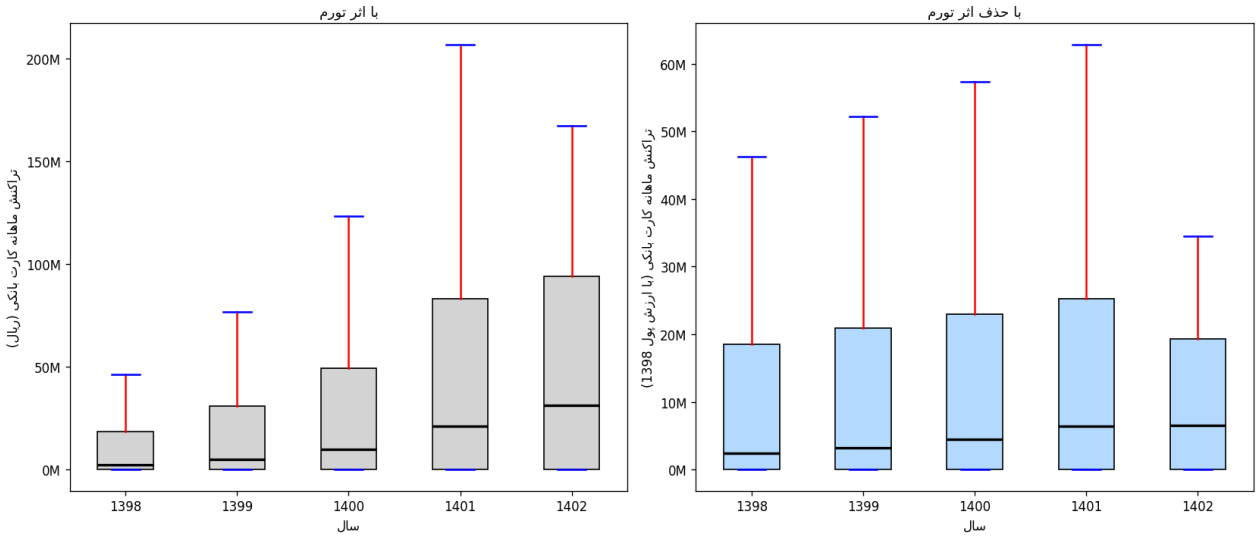
نمودار ۵. فراوانی افراد در دهک‌های درآمدی

هرکدام از دهک‌ها حدود ٪۱۰ از جمعیت را به خود اختصاص داده‌اند که نشان می‌دهد دهک‌بندی جمعیت به خوبی و به‌طور متوازن انجام شده است.



نمودار ۶. میانگین تراکنش کارت بانکی ماهانه در سال ۱۴۰۲ به تفکیک دهک درآمدی
(مقادیر بسیار زیاد و پرت (سه درصد اول)، برای خوانایی نمودار حذف شده‌اند)

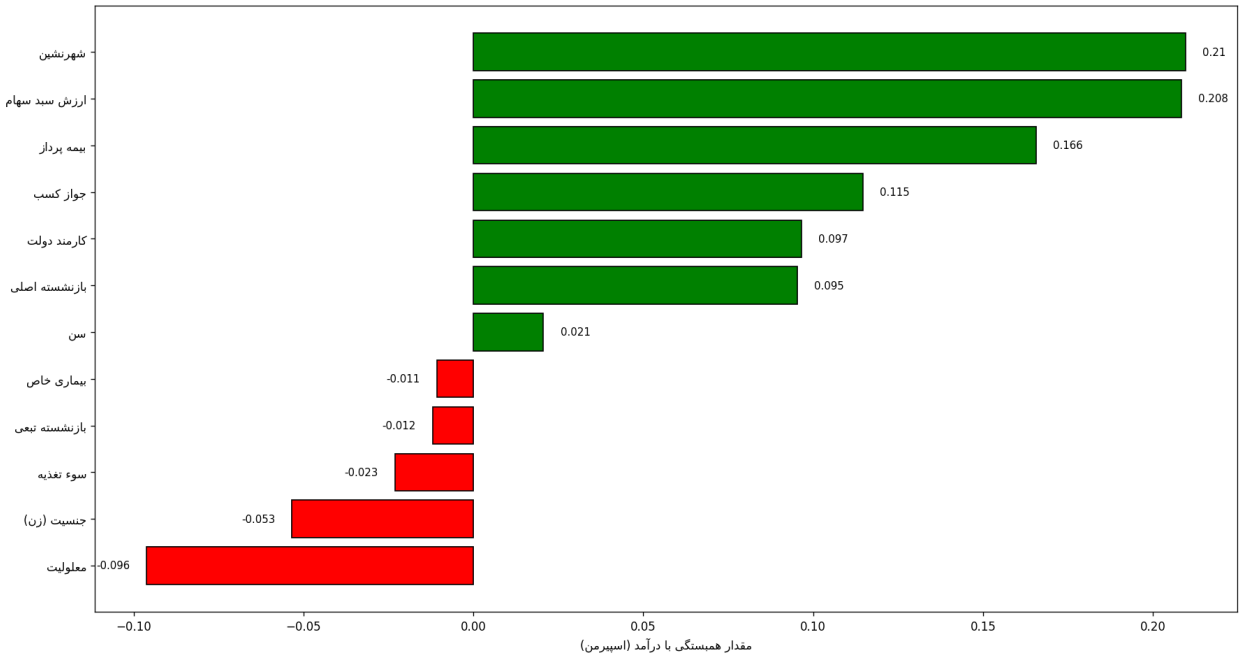
یک الگوی صعودی بسیار قدرتمند میان دهک درآمدی و میانگین تراکنش بانکی افراد مشهود است. میانگین تراکنش‌ها در دهک اول حدود ۱۴ میلیون ریال و در دهک دهم حدود ۴۱۶ میلیون ریال است که نمایانگر جهش ۳۰ برابری این مقدار است. نکتهٔ قابل توجه این است که در دهک اول حتی داده‌های پرت[19] نیز بالاتر از ۳۰۰ میلیون ریال نرفته‌اند.



تراکنش بانکی ماهانه در سال‌های مختلف با حذف و بدون حذف تورم[20]نمودار ۷. مقایسهٔ نمودارهای جعبه‌ای

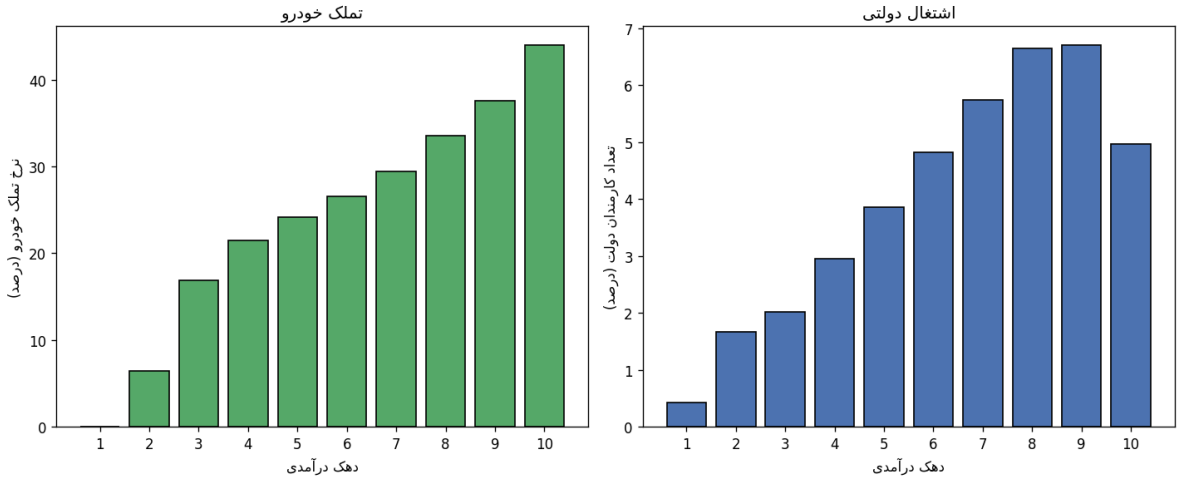
[19] Outlier Data
[20] Box Plot

نمودار ۷ نشان می‌دهد که مقدار ریالی میانگین مبلغ ماهانهٔ تراکنش‌های بانکی در سنوات متوالی، روند افزایشی چشمگیری داشته است. با این حال، بخش اعظم این افزایش را می‌توان به‌علت وجود تورم قابل توجه در این سال‌ها در ایران دانست. به همین سبب، با حذف اثر تورم از تراکنش‌های بانکی (طبق نرخ اعلامی بانک مرکزی) در نمودارهای جعبه‌ای سمت راست، دید متفاوتی نسبت به این موضوع خواهیم داشت.



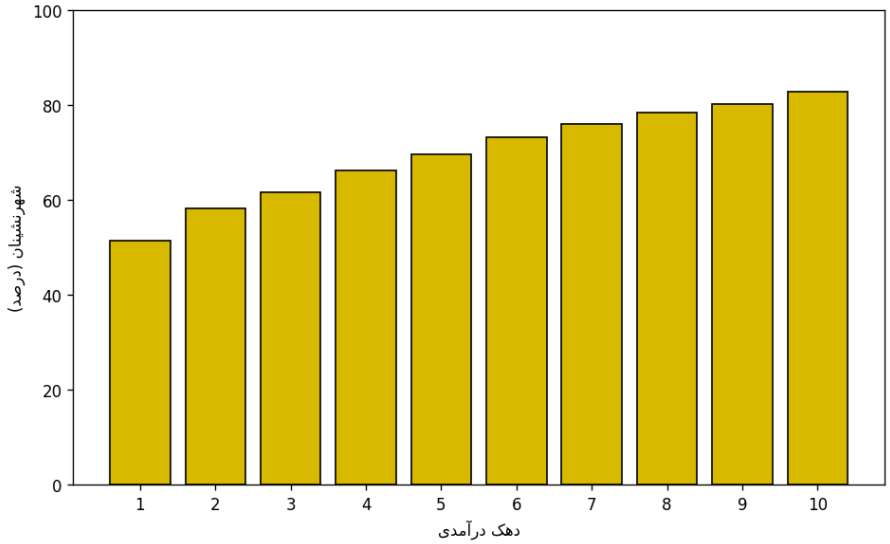
نمودار ۸. همبستگی میان شاخص‌های مختلف با دهک درآمدی

۳.۴ . اشتغال و اموال افراد در هر دهک



نمودار ۹. تملک خودرو و کارمندی دولت به تفکیک دهک‌های درآمدی

طبق آنچه نمودار ۹ نشان می‌دهد، تملک خودرو با افزایش دهک درآمدی در افراد افزایش می‌یابد. به‌طوری که در دهک ۱ تقریباً هیچکس خودرویی ندارد و در دهک دهم، حدود نیمی از افراد حداقل یک خودرو دارند. همچنین سهم کارکنان دولت نیز در دهک‌ها روند صعودی دارد و البته در دهک دهم کاسته می‌شود. مقایسهٔ این دو هیستوگرام در کنار هم با توجه به دهک دهم، نشان می‌دهد اگرچه دهک دهم از وضع مالی بهتری برخوردار است، اما این افراد با اختلاف زیادی نسبت به دهک‌های ۹ و ۸ و ۷، کمتر در مشاغل دولتی مشغول به کار هستند. می‌توان از این مشاهده این برداشت را داشت که کارمندان مشاغل دولتی از درآمدی متوسط برخوردارند و بیش‌از نیمی از افراد ثروتمند (دهک دهم) در مشاغل غیردولتی فعال هستند. از طرفی اکثر جمعیت دهک‌های ۶ تا ۹ کارمند دولت هستند.



نمودار ۱۰. سهم شهرنشینان در دهک‌های درآمدی

شهرنشینان به‌طور قابل توجهی سهم بیشتری در دهک‌های درآمدی بالاتر دارند. این به معنای درآمد بیشتر در مناطق شهری و درآمد کمتر در مناطق روستایی است.

۵. بحث و توضیح نتایج

در این بخش به سوالاتی که در فصل ۲ مطرح شد پاسخ می‌دهیم.

میانگین مبلغ ماهانهٔ تراکنش‌های بانکی در دهک‌های مختلف چه تفاوتی با یکدیگر دارند؟

میانگین تراکنش‌های ماهانه از حدود **۱۴ میلیون ریال در دهک اول** به حدود **۴۱۶ میلیون ریال در دهک دهم** افزایش می‌یابد (نمودار ۶). این الگو نشان‌دهندهٔ این است که طبقه‌بندی دهک‌های درآمدی مورد استفادهٔ دولت، تا حد قابل قبولی با رفتار مالی افراد همخوانی دارد و استفاده از این سیستم برای هدفمندسازی یارانه مناسب به‌نظر می‌رسد.

آیا در طول سال‌های ۱۳۹۸ تا ۱۴۰۲، یک روند صعودی در میانگین تراکنش‌های بانکی مشاهده می‌شود؟

بله. نمودارهای جعبه‌ای رسم شده نشان می‌دهند که تراکنش‌های ماهانهٔ کارت بانکی از ۱۳۹۸ تا ۱۴۰۲، روندی افزایشی دارند. با این حال، در حالی که مقدار ریالی تراکنش‌ها در طول این سال‌ها اکیداً صعودی بوده است، ارزش حقیقی تراکنش‌ها و قدرت خرید افزایش چندانی نیافته است. چرا که پس از حذف اثر تورم از داده‌ها مشاهده می‌شود که در سال‌های ۱۳۹۸ تا ۱۴۰۱ یک روند صعودی ضعیف و پس از آن یک روند نزولی وجود داشته است (نمودار ۷).

کدام‌یک از متغیرها بیشترین ارتباط را با دهک درآمدی اشخاص دارند؟

تمام ضرایب همبستگی به‌دست‌آمده میان متغیرهایی که مستقیماً درآمد افراد را نمایش نمی‌دهند، از نظر قدر مطلق کمتر از ۰.۳ هستند که نشان می‌دهد هیچ‌یک از این متغیرها به‌تنهایی همبستگی قدرتمندی با دهک درآمدی افراد ندارند. با این حال، پنج عامل با بیشترین همبستگی مثبت با دهک درآمدی افراد به‌ترتیب عبارت‌اند از: ۱. شهرنشینی (ρ=+0.210)، ۲. ارزش سبد سهام (ρ=+0.208)، ۳. بیمه‌پرداز بودن (ρ=+0.166)، ۴. دارای جواز کسب (ρ=+0.115) و ۵. کارمندی دولت (ρ=+0.097). همچنین پنج عامل با بیشتری همبستگی منفی با دهک درآمدی افراد به‌ترتیب عبارت‌اند از: ۱. معلولیت (ρ=-0.096)، ۲. جنسیت خانم (ρ=-0.053)، ۳. سوء تغذیه (ρ=-0.023)، ۴. بازنشستهٔ تبعی بودن (ρ=-0.012) و ابتلا به بیماری خاص (ρ=-0.011).

۶. نتیجه‌گیری

پروژهٔ حاضر با استفاده از روش‌های آمار توصیفی در زبان برنامه‌نویسی پایتون، یک مجموعه‌دادهٔ دولتی مربوط به وزارت تعاون، کار و رفاه اجتماعی ایران که شامل حدود ۱.۷ میلیون سطر داده است را به‌منظور بررسی توصیفی و ارتباط میان طبقه‌بندی دهک‌های درآمدی با رفتار مالی افراد، دارایی‌ها و متغیرهای جمعیت‌شناختی تحلیل کرد. نتایج نشان داد که بین دهک درآمدی و مقدار ماهانهٔ تراکنش‌های کارت بانکی یک الگوی صعودی قدرتمند وجود دارد. همچنین ارزش واقعی تراکنش‌های کارت بانکی در طی سال‌های ۱۳۹۸ تا ۱۴۰۲تقریباً ثابت مانده است. ویژگی‌های جمعیت‌شناختی مانند شهرنشینی، تملک خودرو و کارمندی دولت با افزایش دهک درآمدی روندی صعودی نشان دادند. از طرفی میان خصوصیاتی مثل جنسیت، معلولیت، بازنشستگی تبعی و سوء تغذیه همبستگی منفی ضعیفی با دهک درآمدی افراد یافت شد.

۷. بیانیهٔ استفاده از ابزار هوش مصنوعی

در فرایند آماده‌سازی این پروژه، نویسنده از ابزار ChatGPT (OpenAI, GPT-5.5) برای پاسخ به پرسش‌های خود در حوزهٔ مفاهیم آماری (از جمله انحراف معیار، انتخاب ضریب همبستگی مناسب مانند پیرسون یا اسپیرمن، نرخ تورم و سایر مفاهیم مشابه)، رفع خطاهای برنامه‌نویسی و بهبود کدهای تحلیل داده و ترسیم نمودارها (به‌ویژه نمودارهای ۶، ۷ و ۸)، و همچنین دریافت راهنمایی دربارهٔ ساختار پژوهش، شیوهٔ منبع‌دهی، فصل‌بندی، قالب‌بندی و رفع موانع در نرم‌افزار Microsoft Word استفاده کرده است. تمام تحلیل‌ها، کدها، نتایج، نمودارها و متن نهایی پروژه توسط نویسنده به‌طور کامل بازبینی، اصلاح و سپس تأیید شده‌اند. مسئولیت نهایی صحت و اصالت محتوای این اثر بر عهدهٔ پژوهشگر است.

۸. منابع

۱. مدرسهٔ پردازش و تحلیل دادهٔ دقیقه. (۱۴۰۳، ۲۷ خرداد). معرفی داده‌های بانک رفاه ایرانیان. بازیابی شده از: d-learn.ir/wellfare-data-bank

۲. پایگاه اطلاعات رفاه ایرانیان وزارت تعاون، کار و رفاه اجتماعی. (۱۴۰۲). نمونهٔ ۲ درصدی از پایگاه اطلاعات رفاه ایرانیان. بازیابی شده از: دانلودداده–refahdb.mcls.gov.ir/fa/downloaddata

۳. محمدی، هانیه. (۱۴۰۵، ۳۱ اردیبهشت). دریافت یارانه به شرط قبولی در «آزمون وسع». ایبنا. بازیابی شده از: ibena.ir/fa/news/183279

۴. غضنفری‌راد، سعید. (۱۴۰۴، ۹ مهر). پروژهٔ نمونهٔ درس آمار و احتمال مهندسی (sample_project_01.pdf). Github. بازیابی شده از: github.com/saeedgrad/Engineering_Statistics

۵. بانک مرکزی ایران. نرخ تورم. بازیابی شده از: cbi.ir/Inflation/Inflation_fa.aspx

۹. واژگان

English word or terminology	واژه یا اصطلاح به فارسی
Descriptive Statistics	آمار توصیفی
Means Test Algorithm	الگوریتم آزمون وسع
Python	پایتون
Bank Card Transactions	تراکنش‌های کارت بانکی
Income-Decile	دهک درآمدی
Demographics	خصوصیات جمعیت‌شناختی