



شیوه نامه انجام پروژه پایانی

کلاس مبانی برنامه سازی کامپیوتری پایتون
ویژه دانشجویان دانشکده صنایع
دانشگاه صنعتی شریف

علیرضا کدیور
فرحانه جلیلیان، سینا دهباشی، امیرحسین فتحی،
مصطفی قدیمی، علی غلامزاده، حمیدرضا مهبجی

پاییز ۱۳۹۸

- روال انجام کار
- تقویم
- موارد قابل تحویل
- ویژگی های محتوای قابل تحویل
- نمره و معیارهای داوری
- موضوعات

روال انجام کار

اسامی و شماره دانشجویی اعضای گروه **سه یا چهار نفره** مورد نظر خود و یکی از موضوعات را انتخاب و به آدرس kadivar.alireza@gmail.com ایمیل کنید. موضوع انتخابی قابل تغییر نخواهد بود. یکی از افراد حل تمرین به عنوان راهنما برای هر گروه معرفی می‌شود تا اگر در طول مدت انجام پروژه سوالی داشتید با ایشان در تماس باشید.

اجزای مختلف منبع داده مورد نظر خود را دانلود و بدون این که تغییری را به صورت دستی (manual) در فایل‌های منبع ایجاد کنید با استفاده از کد پایتون آن را بصورت یک جدول واحد و با ساختار مناسب یکدیگر بچسبانید و با فرمت csv ذخیره کنید. سپس به پرسش‌ها و درخواست‌های مطرح شده پاسخ دهید. پس از تحویل نهایی فایل‌ها با تعیین وقت قبلی در یک جلسه حضوری کار خود را ارائه می‌دهید. گزارش نهایی شما در روی کوئرا بارگذاری خواهد شد و برای سایر دانشجویان قابل رویت خواهد بود.

تقویم

- انتخاب گروه و موضوع دوشنبه ۱۸ آذر ۱۳۹۸
- تخصیص یک راهنما برای هر گروه سه‌شنبه ۱۹ آذر ۱۳۹۸
- ارسال موارد قابل تحویل پروژه دوشنبه ۱۴ بهمن ۱۳۹۸
- ارائه حضوری پروژه شنبه ۱۹ بهمن و یکشنبه ۲۰ بهمن ۱۳۹۸

موارد قابل تحویل

خروجی قابل تحویل شما شامل موارد زیر می‌شود:

۱. مجموعه کدهایی که به زبان پایتون نوشته‌اید (فایل‌های .py).
۲. جدول تجمیع شده‌ای که تهیه کرده‌اید (فایل csv) و سایر منابع داده‌ای که در تحلیل خود از آن‌ها استفاده کرده‌اید (اگر حجم آن زیاد و غیر قابل ارسال بود بخشی از آن را ارسال کنید).
۳. گزارش و تحلیل (فایل pdf)

ویژگی‌های محتوای قابل تحویل

فایل pdf گزارش و تحلیل شما باید موارد زیر را شامل شود:

- a. **گزارش تقسیم کار:** در این بخش شرح داده می‌شود تقسیم کار بین اعضای تیم چگونه انجام شده و هر کسی کدام بخش از پروژه را انجام داده است.
- b. **توصیف داده:** ساختار جدول یکپارچه‌ای که تهیه کردید را گزارش کنید و هر ستون آن را به اختصار شرح دهید (داده‌ای برای تحلیل آماده کرده‌اید باید ساختاری مناسب برای پاسخ به پرسش‌های تحلیلی مورد نظر داشته باشد).
- c. **توضیح کدهای پایتون:** توضیح مختصری در باره کدهایی که به زبان پایتون نوشته‌اید. این توضیح لازم است بگوید هر فایل py چه ورودی می‌گیرد و چه خروجی می‌دهد. کدنویسی شما باید modular باشد و هر فایل py تا حد امکان یک کار بیشتر انجام ندهد. مثلاً فایل مربوط به جمع داده با فایل تحلیل و پاسخ به پرسش جدا باشد.
- d. **قابلیت به روز رسانی:** در این قسمت باید روشی ارائه شود که در صورت به روز شدن منابع داده با چه ترتیبی و چگونه با اجرای فایل‌های py خروجی‌های پروژه را با داده‌های خام ورودی جدید به روز کنیم. ترسیم دیاگرام برای فهم بهتر کمک می‌کند. نیازی به برنامه‌نویسی برای به روزرسانی تمام خودکار در این پروژه نیست.
- e. **گزارش توصیفی از اطلاعات جمع کلان:** گزارشی توصیفی از اطلاعات کلانی که می‌توان از داده‌ی جمع شده حول موضوع مورد نظر به دست آورد. هدف این است که مخاطب شما متوجه شود مجموعاً در باره چه تعداد رکورد، چه بازه زمانی، چند ریال و ... حرف زده می‌شود.
- f. **کیفیت سنجی:** کیفیت داده‌ای که جمع کرده‌اید را چگونه ارزیابی می‌کنید؟ آیا داده‌ی جمع شده کیفیت مناسبی برای پاسخ‌گویی به پرسش مطرح شده را دارد؟ نتیجه بررسی‌هایی که برای کیفیت سنجی داده انجام می‌دهید در هر دو صورت مشاهده سازگاری یا ناسازگاری گزارش کنید.
- g. **گزارش تحلیلی:** پاسخ به پرسش‌های مطرح شده در باره منبع داده‌ای که انتخاب کرده‌اید.
- h. **طرح پرسش:** علاوه بر پرسش‌ها و درخواست‌های مطرح شده برای هر موضوع یک پرسش که ارزش تحلیلی داشته باشد و بتوان با استفاده از داده‌های در دسترس به آن پاسخ داد طرح کنید و به آن پاسخ دهید.
- i. استفاده از **مصورسازی‌های گویا و خلاصه، جدول و نمودار و دیاگرام** در انتقال مطلب به شما کمک می‌کند و استفاده از آن‌ها اکیداً توصیه می‌شود اما در استفاده از آن زیاده روی نکنید.
- j. رنگ باید تا جای امکان کارکرد مشخصی در گزارش شما داشته باشد. سعی کنید از رنگ به صورت غیرهدفمند استفاده نکنید.
- k. قسمتی به نام **مقدمه** در کمتر از یک صفحه در ابتدای گزارش خود بیاورید و در آن بگویید این گزارش به صورت کلی حاوی چه بخش‌ها و مطالبی می‌شود. علاوه بر آن بخشی از دستاوردهای کار خود که به نظرتان ارزش اطلاعاتی و تحلیلی بیشتری دارد را به صورتی در ابتدای گزارش بیاورید تا مخاطب انگیزه پیدا کند تمام آن را مطالعه کند.

نمره

چنانچه در ابتدای ترم مطرح شده بود نمره پروژه ۱۸٪ از نمره کل درس است. اما در صورت تحویل عالی کار این نمره تا سقف ۳۶٪ و ۳۲٪ به ترتیب برای گروه‌های سه نفره و چهار نفره قابل افزایش است. اگر در بخش‌های دیگر مانند میان‌ترم نمره از دست داده‌اید می‌توانید با صرف زمان روی پروژه خود و ارائه کار با کیفیت آن را جبران کنید.

معیارهای داوری

۱. **تقسیم کار** فکر شده و متعادل بین اعضای گروه (به عنوان مثال یک نمونه تقسیم فکر نشده تقسیم کاری است که در آن دوباره‌کاری وجود دارد)
۲. **ارتباط با راهنمای پروژه** در طول مدت انجام کار و در جریان گذاشتن ایشان از روند پیشرفت کار و بیان مشکلات و ابهام‌ها
۳. **وقت‌شناسی** با توجه به تقویم مطرح شده
۴. **توصیف داده** به صورتی که مخاطب بتواند به راحتی متوجه شود داده‌های در دسترس چیست و حاوی چه اطلاعاتی می‌شود
۵. اتخاذ استراتژی مناسب برای **تقسیم‌بندی فایل‌های پایتون**
۶. **کیفیت کدنویسی** مانند کامنت‌گذاری کدها، نام‌گذاری مناسب و قابل فهم متغیرها و ...
۷. ارائه توضیح برای کدها
۸. **قابلیت به روزرسانی** نیمه خودکار در صورت دریافت داده‌ی جدید با همان ساختار
۹. **گزارش توصیفی** از اطلاعات سرجمع و کلان
۱۰. **کیفیت سنجی**
۱۱. **پاسخ به پرسش‌ها**
۱۲. **کیفیت استدلال‌ها**
۱۳. **سازگاری و صحت نتایج**
۱۴. **طرح پرسش و پاسخ به آن**
۱۵. استفاده مناسب و به اندازه از **رنگ و مصورسازی** اطلاعات مانند نمودار و ...
۱۶. ارائه توضیح کافی برای **فهم مستقل نمودار و تصویر** در قالب زیرنویس یا محتوای داخل نمودار
۱۷. **ادبیات موجز و مناسب**
۱۸. **قدرت استدلال و تحلیل**
۱۹. آیا کار در ارائه دستاوردهای اصلی خود به زبان ساده و روان و مشخص و ایجاد انگیزه برای پیگیری موضوع در مخاطب موفق بوده است؟
۲۰. **ارائه حضوری**

موضوعات

۱. تحلیل توزیع کودکان بازمانده از تحصیل در شهرستان‌های کشور

مرکز آمار در هر سرشماری جمعیت کل و جمعیت در حال تحصیل را به تفکیک سن، جنس و وضعیت سکونت (شهری، روستای و غیرساکن) در شهرستان‌های کشور منتشر می‌کند. تفاضل این دو جمعیت، شامل کسانی خواهند بود که به مدرسه نمی‌روند. اگر این جمعیت در سن مدرسه (مثلاً ۷ تا ۱۵ سال) قرار داشته باشد جمعیت بازمانده از تحصیل محسوب می‌شود.

پرسش‌ها و درخواست‌ها

- با استفاده از داده‌های در دسترس تحلیلی از وضعیت بازماندن از تحصیل در سال‌های ۹۰ و ۹۵ در کشور ارائه دهید:
- توزیع درصد بازماندن از تحصیل را در سال‌های ۹۰ و ۹۵ به تفکیک جنس، وضعیت سکونت و گروه‌های سنی ۷ تا ۱۰، ۱۱ تا ۱۴ و ۱۵ تا ۱۸ ترسیم و تفسیر کنید.
- درصد بازماندن از تحصیل از سال ۹۰ تا ۹۵ در کدام یک از شهرستان‌های کشور (با حفظ تفکیک‌های بالا) بیشترین افزایش و بیشترین کاهش را داشته است؟

منبع داده: [نتایج سرشماری مرکز آمار ایران](#) (اطلاعات جمعیتی و تحصیلی)

۲. تحلیل مهاجرت بین شهرستان‌های کشور از ۱۳۹۰ به ۱۳۹۵

مرکز آمار جمعیت مهاجر بین شهرستان‌های کشور در ۵ سال منتهی به سرشماری را در قالب ماتریس مهاجرت منتشر می‌کند. در این ماتریس مشخص می‌شود که در این بازه زمانی چه جمعیتی از کدام شهرستان به کدام شهرستان مهاجرت کرده است.

پرسش‌ها و درخواست‌ها

- تحلیلی از مهاجرت بین شهرستان‌های کشور ارائه کنید:
- توزیع درصد جمعیت مهاجر در شهرستان‌های کشور در پنج سال منتهی به سال ۹۰ و ۹۵ را ترسیم و تفسیر کنید. آیا تغییری در این توزیع رخ داده است؟
- در ۵ سال منتهی به ۹۵ نسبت به ۵ سال منتهی به ۹۰ مهاجرت از کدام شهرستان‌ها به کدام شهرستان‌ها بیشترین افزایش و بیشترین کاهش را داشته است؟
- مهاجرپذیرترین و مهاجرپذیرترین شهرستان‌ها نسبت به جمعیت خود کدام شهرستان‌ها بوده‌اند؟
- الگوی جابجایی بین شهرستان‌های کشور چگونه است؟

منبع داده: [نتایج سرشماری مرکز آمار ایران](#) (ماتریس‌های مهاجرت و اطلاعات جمعیتی)

۳. اطلاعات مدیران و کارکنان شهرداری تهران

سامانه شفافیت شهرداری تهران اطلاعات مدیران و کارکنان شهرداری تهران را این سامانه منتشر کرده است. این اطلاعات وضعیت استخدام، سابقه کار، محل خدمت، مرتبه، سمت، تاریخ انتصاب، محل و رشته و آخرین مقطع تحصیل و سمت‌های قبلی فرد را شامل می‌شود.

پرسش‌ها و درخواست‌ها

تحلیلی از آمار کارکنان شهرداری و میزان پرداخت آن‌ها ارائه کنید:

- تعداد کارکنان را به تفکیک وضعیت استخدام، محل خدمت، مقطع تحصیلی، رشته تحصیلی، مرکز آموزشی محل تحصیل در قالب جدول و به صورت بصری ترسیم کنید.
- توزیع مبلغ حکم را به تفکیک وضعیت استخدام، محل خدمت، مقطع تحصیلی، رشته تحصیلی، مرکز آموزشی محل تحصیل ترسیم کنید. آیا تمایز چشمگیری در وضعیت میزان پرداخت (به تفکیک حالت‌های فوق) می‌بینید؟
- توزیع پرداخت در مناطق مختلف (محل خدمت) را مقایسه کنید. آیا تمایز چشمگیری بین آن‌ها وجود دارد؟
- آیا اصل برابری در حکم حقوق در مناطق مختلف شهرداری کم و بیش رعایت شده است؟ به بیان دیگر آیا همبستگی میان میانگین پرداخت به پرسنل در مناطق مختلف و قیمت متر مربع مسکن در آن مناطق وجود دارد؟ نمودار پراکندگی نقطه‌ای میانگین پرداخت و قیمت متر مربع منطقه را رسم کنید.

منبع داده: [سامانه شفافیت شهرداری تهران](#) (اطلاعات مدیران ارشد، فهرست مشخصات کارکنان)؛ [تهران هوشمند، داده‌های تهران](#) (ساختمان و مسکن، املاک، لیست داده‌ها)

۴. قراردادهای شهرداری تهران

سامانه شفافیت شهرداری تهران اطلاعات قراردادهای هزینه‌ای، درآمدی و مشارکتی-سرمایه‌گذاری شهرداری تهران را به صورت مستمر منتشر می‌کند (توضیحات داده در صفحات مربوطه آمده است). این اطلاعات تاریخ امضا، تاریخ شروع، تاریخ پایان، مبلغ قرارداد، نام و شناسه ملی و کد اقتصادی پیمانکار، روش انتخاب پیمانکار، نام کارفرما و موضوع قرارداد را شامل می‌شود.

پرسش‌ها و درخواست‌ها

- با استفاده از داده‌های در دسترس تحلیلی از وضعیت قراردادهای شهرداری تهران ارائه کنید:
- تعداد و مجموع مبلغ و متوسط مبلغ قراردادهای هزینه‌ای، درآمدی و مشارکتی-سرمایه‌گذاری برای قراردادهایی که مبلغ بیشتر از صفر برای آن‌ها درج شده چقدر است؟

- توزیع مبلغ قراردادهای دارای مبلغ بیشتر از صفر را به تفکیک روش انتخاب پیمانکار ترسیم و با یکدیگر مقایسه کنید. آیا تمایز قابل توجهی بین آنها وجود دارد؟
- مجموع مبلغ قرارداد پیمانکارهای مختلف را محاسبه و توزیع آن را ترسیم و تفسیر کنید. آیا می‌توانید پیمانکارانی که مجموع مبلغ قراردادهای آنها تمایزی چشمگیر با سایرین دارد را مشخص کنید؟ مرز ریالی تعیین پیمانکاران متمایز را چگونه تعیین می‌کنید؟
- نسبت مجموع مبلغ ترک تشریفات به کل مبلغ قرارداد را برای پیمانکاران و کارفرمایان مختلف محاسبه و توزیع آن را ترسیم و تفسیر کنید. آیا می‌توانید پیمانکاران و کارفرمایانی که این نسبت در آنها تمایزی چشمگیر با سایرین دارد را مشخص کنید؟ مرز مقداری تعیین آنها را چگونه تعیین می‌کنید؟
- آیا ارتباطی از جنس همبستگی بین مجموع مبلغ قرارداد و نسبت مجموع مبلغ ترک تشریفات به کل مبلغ قرارداد پیمانکاران وجود دارد؟
- فرض کنید مبلغ قرارداد ارزشی است که شهرداری برای هر کلمه در موضوع قرارداد قائل شده است. نسبت مجموع ارزش ریالی کلمات «سبز»، «پارکینگ»، «بزرگراه»، «پسماند»، «میوه» و «ورزش» به ارزش ریالی کل قراردادها را به صورت ماهانه محاسبه کنید و منحنی سری زمانی آن را ترسیم کنید. آیا این نسبت در طول زمان کم و بیش ثابت است؟ الگوهای تکرارشونده و فصلی دارد؟ هیچ الگوی مشخصی ندارد؟
- انتظار می‌رود نسبت مبلغ قراردادهای دارای موضوع مشابه به طول مدت قرارداد به یکدیگر نزدیک باشند. درستی این گزاره را تحقیق کنید و نمونه‌ای از موضوعاتی که قیمت مشابه دارند و موضوعاتی که قرارداد مشابه ندارند را گزارش کنید. آیا می‌توان در موضوعاتی که قیمت واحد زمان کم و بیش مشابه دارند ناهمخوانی قیمت در قراردادهای مختلف را استخراج کرد؟
- با توجه به نام کارفرما مانند «شهرداری منطقه ۱۳» قراردادهایی که می‌توان آنها را به یکی از مناطق ۲۲ گانه تهران نسبت مشخص و مجموع مبالغ آنها را به تفکیک مناطق ۲۲ گانه محاسبه کنید. به بیان دیگر محاسبه کنید شهرداری در هر منطقه چقدر هزینه کرده است.
- آیا ارتباطی از جنس همبستگی بین هزینه کرد شهرداری در هر منطقه و قیمت متر مربع زمین در آن منطقه وجود دارد؟ این ارتباط به عدالت اجتماعی نزدیک است یا از آن دور است؟

منبع داده: سامانه شفافیت شهرداری تهران (اطلاعات قراردادها)؛

تهران هوشمند، داده‌های تهران (ساختمان و مسکن، املاک، لیست داده‌ها)

۵. حوادث آب در تهران

داده‌ها مربوط به حوادث آب تهران از آبان تا بهمن ۹۴ با ذکر نوع حادثه، زمان گزارش، زمان اعزام به محل، زمان شروع تعمیر و زمان پایان رسیدگی و سایر اطلاعات مربوط به آن ثبت شده است.

پرسش‌ها و درخواست‌ها

- با استفاده از داده‌های در دسترس، تحلیلی از اطلاعات حوادث آب در تهران ارائه کنید:
- آیا حادثه یا علت بروز حادثه‌ای هست که نسبت رخ دادن آن به کل حوادث، در طول هفته‌های موجود در داده مربوطه افزایش یا کاهش چشمگیر یافته باشد؟
- فاصله زمانی گزارش، اعزام، شروع و پایان فرایند رسیدگی به حوادث را محاسبه کنید. توزیع آن‌ها را ترسیم و با یکدیگر مقایسه کنید.
- آیا این فواصل زمانی در مناطق، عوامل ایجاد، نوع حادثه و اهمیت‌های مختلف با یکدیگر تمایز قابل توجهی دارند؟
- عوامل موثر بر فاصله زمانی شروع تا پایان عملیات رسیدگی را چگونه ارزیابی می‌کنید؟ به نظر شما رسیدگی به چه حوادثی بیشتر و رسیدگی به چه حوادثی کمتر طول می‌کشد؟

منبع داده: آب و فاضلاب تهران (فایل‌های مربوطه در کوئرا بارگذاری شده است).

۶. مصرف برق در تهران

مصرف برق تهران از سال ۱۳۹۰ تا آذر ۱۳۹۷ در هر ساعت در دو فایل آمده است. علاوه بر آن، تغییرات دمای تهران در هر سه ساعت در فایل دیگری آورده شده است.

پرسش‌ها و درخواست‌ها

- با استفاده از داده‌های در دسترس، تحلیلی از مصرف برق تهران ارائه کنید:
- توزیع مصرف برق در روزهای مختلف هفته را رسم کنید. آیا تفاوتی در مصرف برق در روزهای مختلف هفته وجود دارد؟
- توزیع مصرف برق در ساعات مختلف شبانه‌روز را رسم کنید. چه تفاوتی در مصرف برق بین ساعات مختلف شبانه‌روز وجود دارد؟
- توزیع مصرف برق در فصل‌های مختلف سال را رسم کنید. چه تفاوتی در مصرف برق در فصل‌های مختلف سال وجود دارد؟
- رابطه مصرف برق با دما چگونه است؟ آیا ارتباطی از جنس همبستگی بین مصرف برق و دما وجود دارد؟
- به نظر شما کدام یک از عوامل فوق بیشترین تاثیر در تعیین میزان مصرف برق دارد؟
- منحنی مصرف برق در طول ۷ سال را ترسیم کنید. مصرف برق در طول این مدت به طور متوسط چقدر افزایش یافته است؟

منبع داده: برق تهران، سازمان هواشناسی (فایل‌های مربوطه در کوئرا بارگذاری شده است).

۷. اطلاعات پرواز

این داده‌ها برای استفاده دانشجویان، در پروژه درس «مبانی برنامه‌نویسی» اختیار شما قرار گرفته است. از اینکه آن را برای دیگران ارسال یا فوروارد نمی‌کنید از شما سپاسگزاریم. توضیح داده‌ها:

Id: شماره سطر رکورد

User id: شناسه کاربری مشتریان

Request-date-id: تاریخ خرید

Request-time: ساعت خرید

Departure-date-id: ساعت بلیت

Company: کد ایرلاین

Source: کد مبدا

Destination: کد مقصد

قیمت (قیمت بلیت‌ها با قیمت واقعی متفاوت است، به همین دلیل در صورت استفاده از اطلاعات خارجی Price: به این موضوع توجه داشته باشید)

پرسش‌ها و درخواست‌ها

- با استفاده از داده‌های در دسترس، تحلیلی از این داده ارائه کنید:
- با کمک دیتای در دسترس، محاسبه کنید که به طور متوسط هر مشتری چند تومان خرید کرده است.
- معیاری از خرید (مثلاً بیشتر از ۹۰۰ هزار تومان) در نظر بگیرید و برای آن مقادیری به عنوان امتیاز برای محاسبه‌ی امتیازهای هر کاربر در باشگاه مشتریان محاسبه کنید، وقادارترین مشتریان چه کسانی هستند؟
- وضعیت فروش را در روزهای مختلف هفته و ساعت‌های مختلف یک روز بررسی کنید. پرترفدارترین و کم‌ترفدارترین زمان‌ها و روزهای هفته برای فروش کدام هستند؟ آیا مقدار فروش ماهیانه نسبت به ماه حساسیت دارد؟ (برای این کار تاریخ میلادی را به روز تبدیل می‌کنید)
- فاصله‌ی بین خرید و تاریخ بلیت معمولاً نمود خوبی از رفتار مشتری و همچنین اطمینان از سطح سرویس تامین‌کننده را به نمایش می‌گذارد. فاصله‌ی بین خرید و زمان حرکت را در ماه‌های مختلف بررسی کنید و آن را نمایش دهید. ازین مقایسه چه نتیجه‌ای می‌گیرید؟
- هنگام برنامه‌ریزی ایرلاین‌ها برای تخصیص منابع، یکی از موضوعات پراهمیت بررسی استقبال از مسیرهای موجود است. منظور از مسیر دوتایی مبدا-مقصد است. پرترفدارترین مسیرها کدام هستند؟ بین پرترفدارترین مسیرها و پرترفدارترین مبدا و مقصدها چه رابطه‌ای وجود دارد؟
- در خصوص عملکرد ایرلاین‌ها بگویید پرترفدارترین ایرلاین و پردآمدترین ایرلاین کدام هستند؟ وابستگی این موضوع به ماه‌های مختلف و مبدا یا مقصد را بررسی کنید.

پسورد: BXR1fXz@LOM۲