

بوت کیمپ  
تحلیل داده

DATA  
ANALYSIS  
BOOTCAMP



RAHNEMA  
COLLEGE

# بررسی داده‌های تاکسی شهر نیویورک

گروه NEMO

## سوال ۱:

یک خلاصه‌ی آماری از فیلدهای مختلف موجود در داده ارائه کنید.



+۱۴۷,۰۰۰,۰۰۰

سفر



۲۰۲۰ - ۲۰۲۴

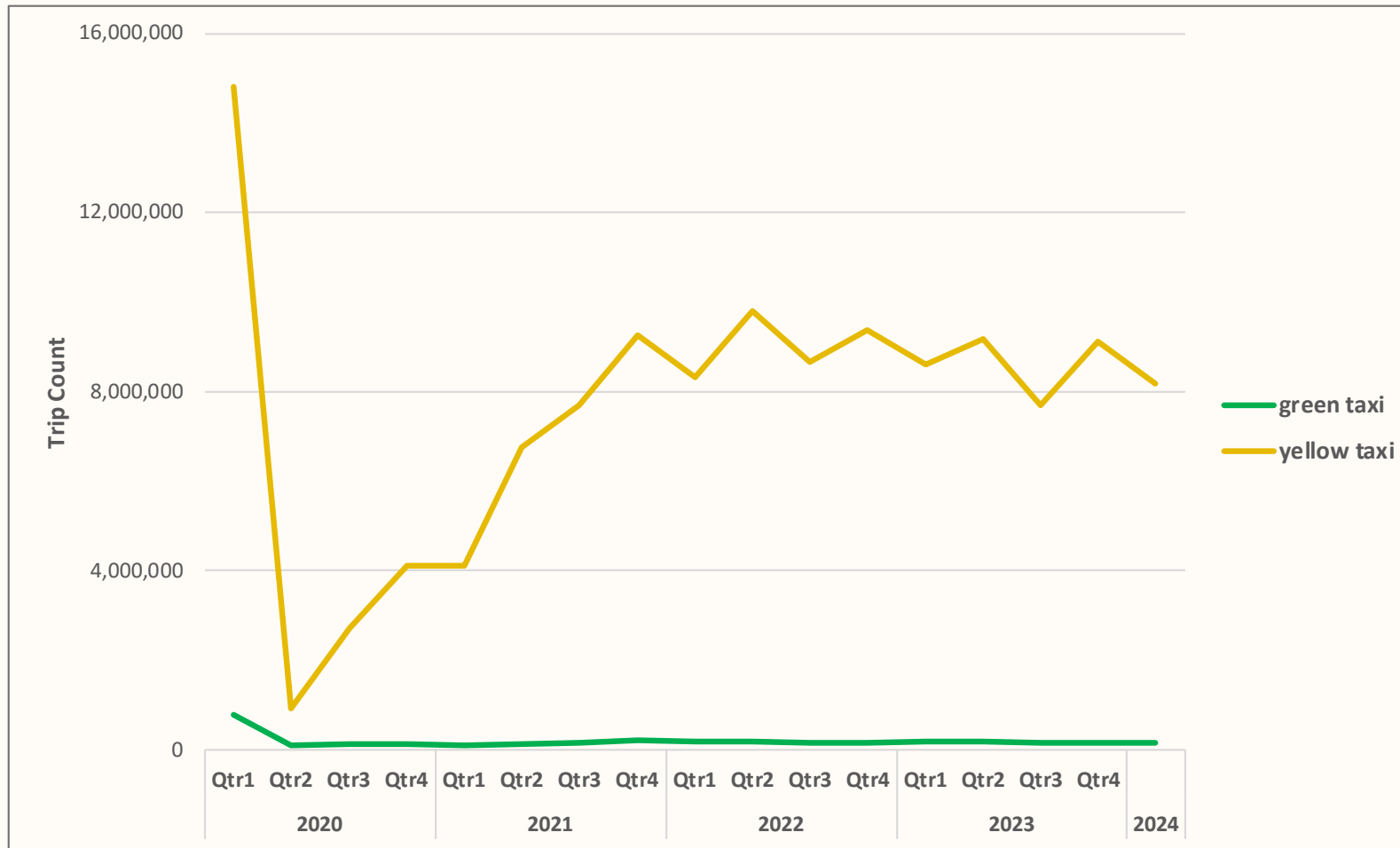
بازه زمانی



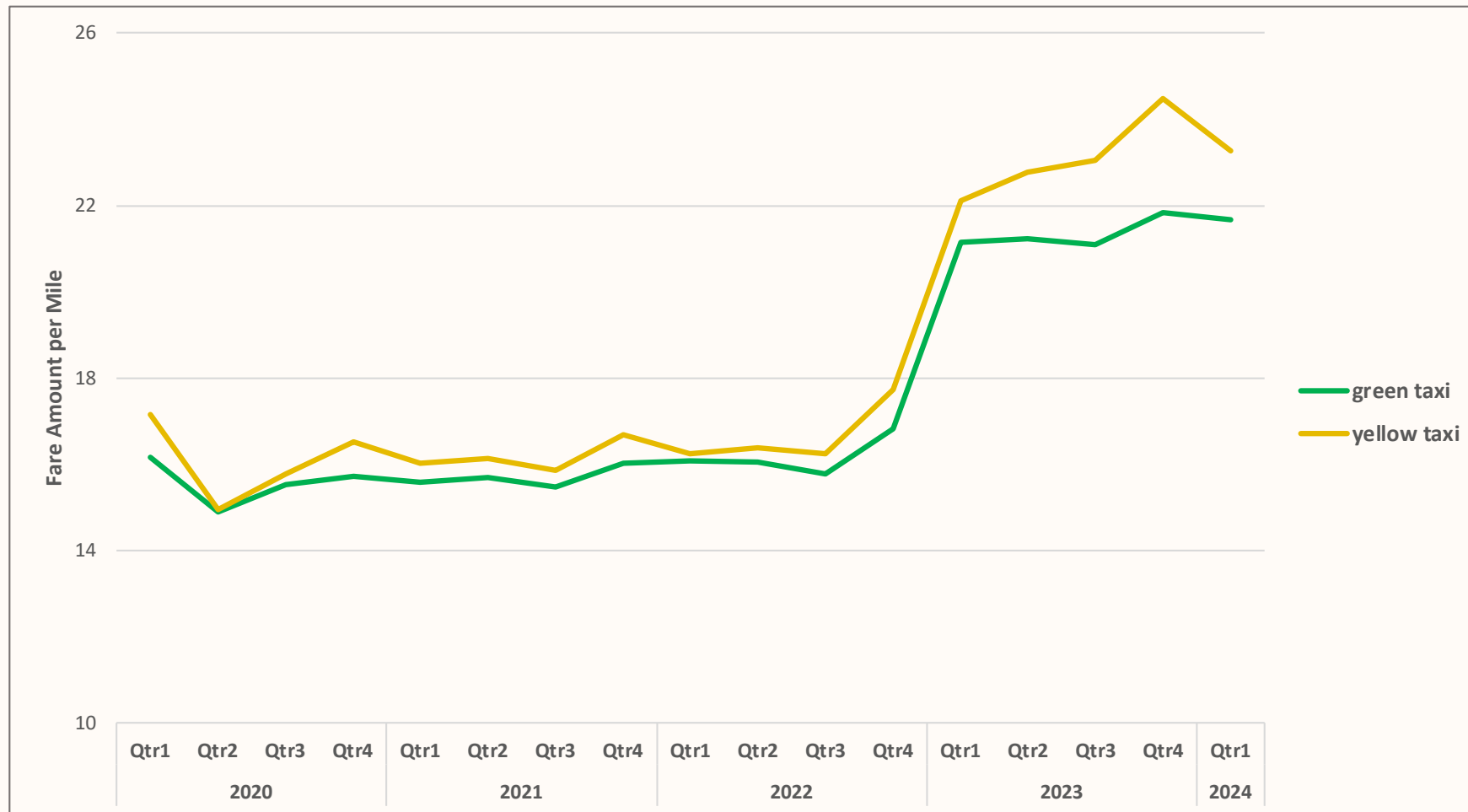
۲,۹۴۳

میلیون دلار

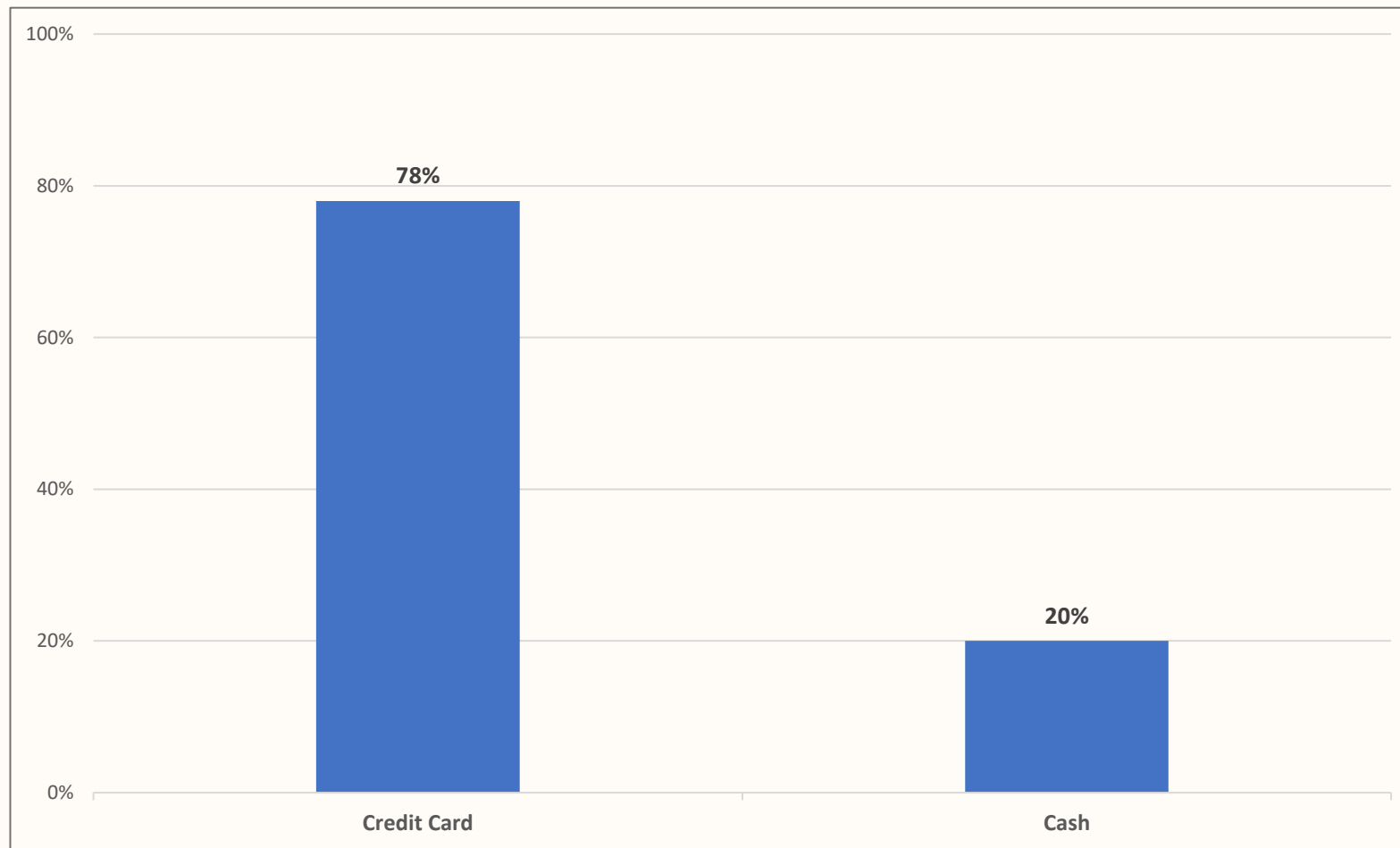
نمودار تعداد سفر در طول ۴ سال گذشته:



نمودار مقدار کرایه به ازای هر مایل در طول ۴ سال گذشته:



## نمودار توزیع نحوه پرداخت مسافران:



## سوال ۲:

گروهی از دوستان شما قصد دارند محصول پرداخت اینترنتی را برای تاکسی‌های زرد و سبز سطح شهر ارائه دهند. به این صورت که هر راننده یک حساب کاربری در اپلیکیشن ایجاد کند و کد آن را به مسافران ارائه کند. مسافران نیز بتوانند یک کیف پول در حساب کاربری خود داشته باشند و هزینه تاکسی را از طریق کد راننده پرداخت کنند.

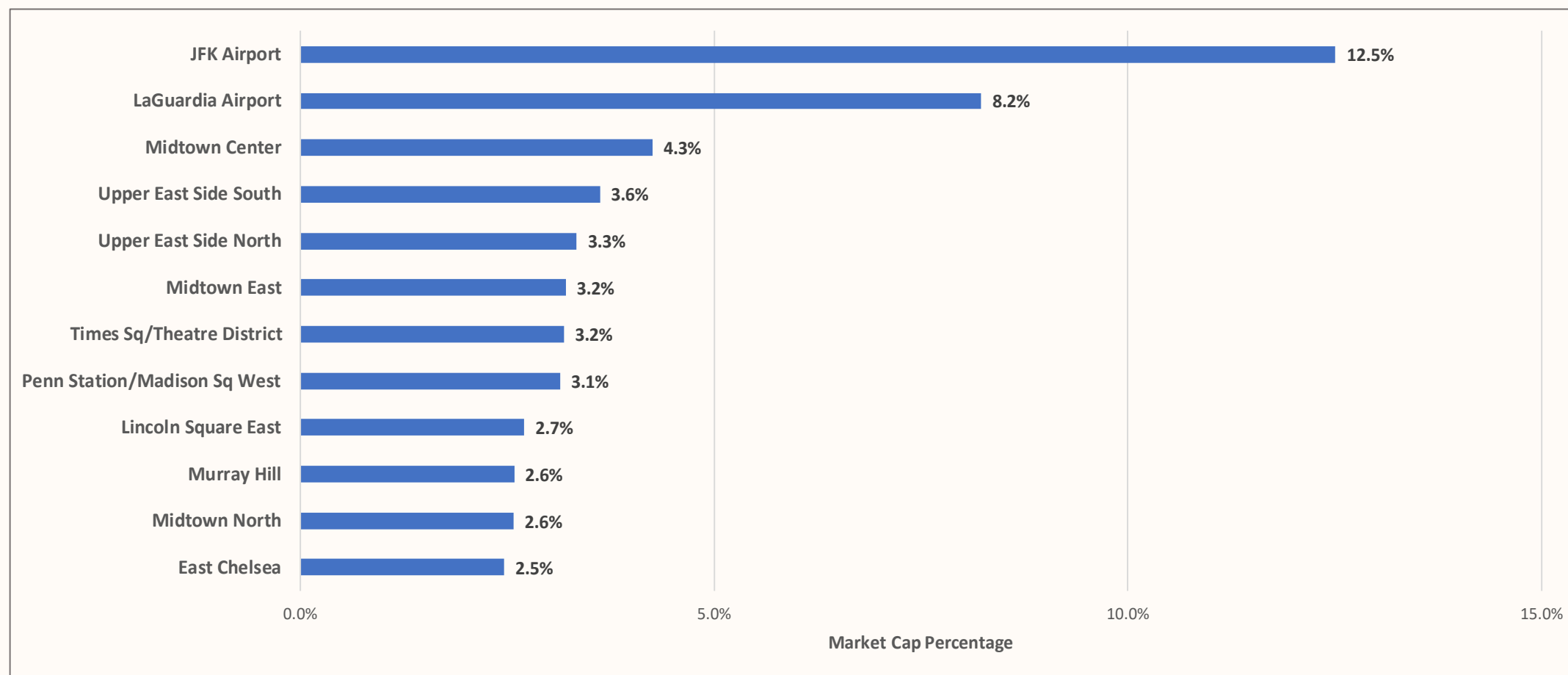
برای راه‌اندازی اولیه محصول چه مسیرهایی را پیشنهاد می‌کنید؟ به نظر شما محصول BNPL (الان بخر، بعداً پرداخت کن) چه قدر می‌تواند سودآوری ایجاد کند؟

- بازه زمانی: ۱۲ ماه اخیر
- پرداخت‌های با Credit Card
- مبداهایی با تعداد و مبلغ تراکنش بالا
- متریک تعریف مسیرهای پیشنهادی:

(شمار سفرها × میانه مبلغ سفرها)

- بررسی مکان‌هایی با مجموع حجم بازار ۵۰٪

نمودار حجم بازار ۱۲ مکان منتخب که ۵۰٪ از حجم بازار را اشغال می‌کنند:

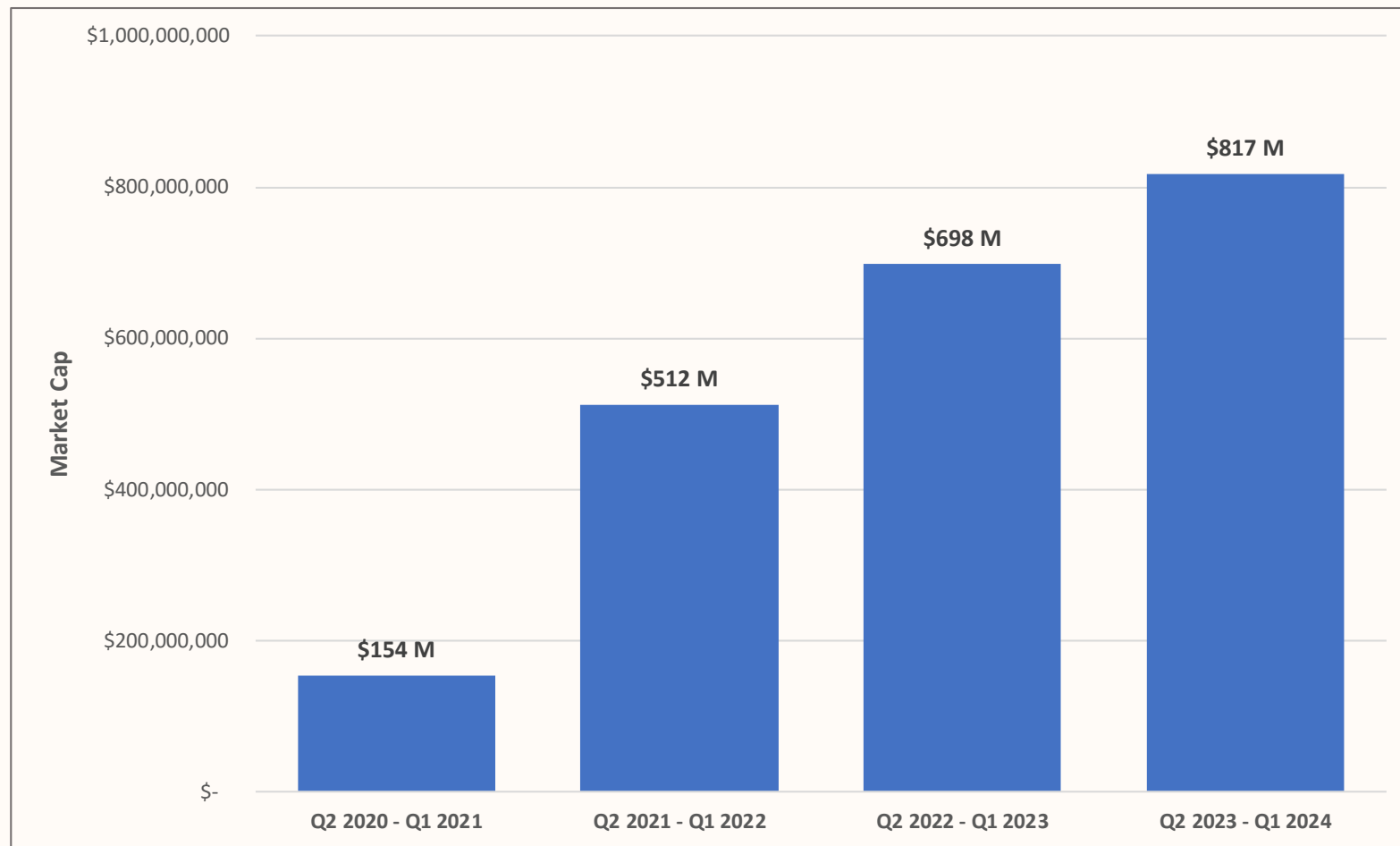




### پتانسیل سودآوری:

- نرخ رشد مرکب سالانه ۷۴٪
- تارگت ۵۰٪ درصد حجم بازار
- با هدف تصاحب ۱۰٪ از تارگت
- بهره‌ی ۱.۵٪ سرویس BNPL

نمودار حجم بازار در طول ۴ سال اخیر:



پتانسیل سودآوری یک ساله:

۱.۱ میلیون دلار

### سوال ۳:

یک شرکت تبلیغاتی قصد دارد از ظرفیت تاکسی‌های سطح شهر به‌عنوان جایگاه تبلیغاتی استفاده کند. این شرکت از شما در مورد پتانسیل تبلیغاتی در این روش و همچنین تاکسی‌های مناسب برای اجرای این برنامه مشاوره خواسته است.

**الف-** با فرضیات معقول و تعریف متریک مشخص کنید این نوع تبلیغات چه قدر می‌تواند پتانسیل درآمدزایی داشته باشد.

**ب-** شرکت تبلیغاتی می‌خواهد در چند مکان مشخص نصب این جایگاه تبلیغاتی را برای تاکسی‌ها پروموت کند. پتانسیل درآمدزایی تاکسی‌های هر مکان چه قدر است؟

• تعریف متریک:

### (طول زمان سفر)

• بازه زمانی: ۱۲ ماه اخیر

• CPM معادل \$۱.۷۶

• میانگین زمان ویدیو تبلیغاتی معادل ۳۰ ثانیه

• سهم آژانس تبلیغات معادل ۱۵٪

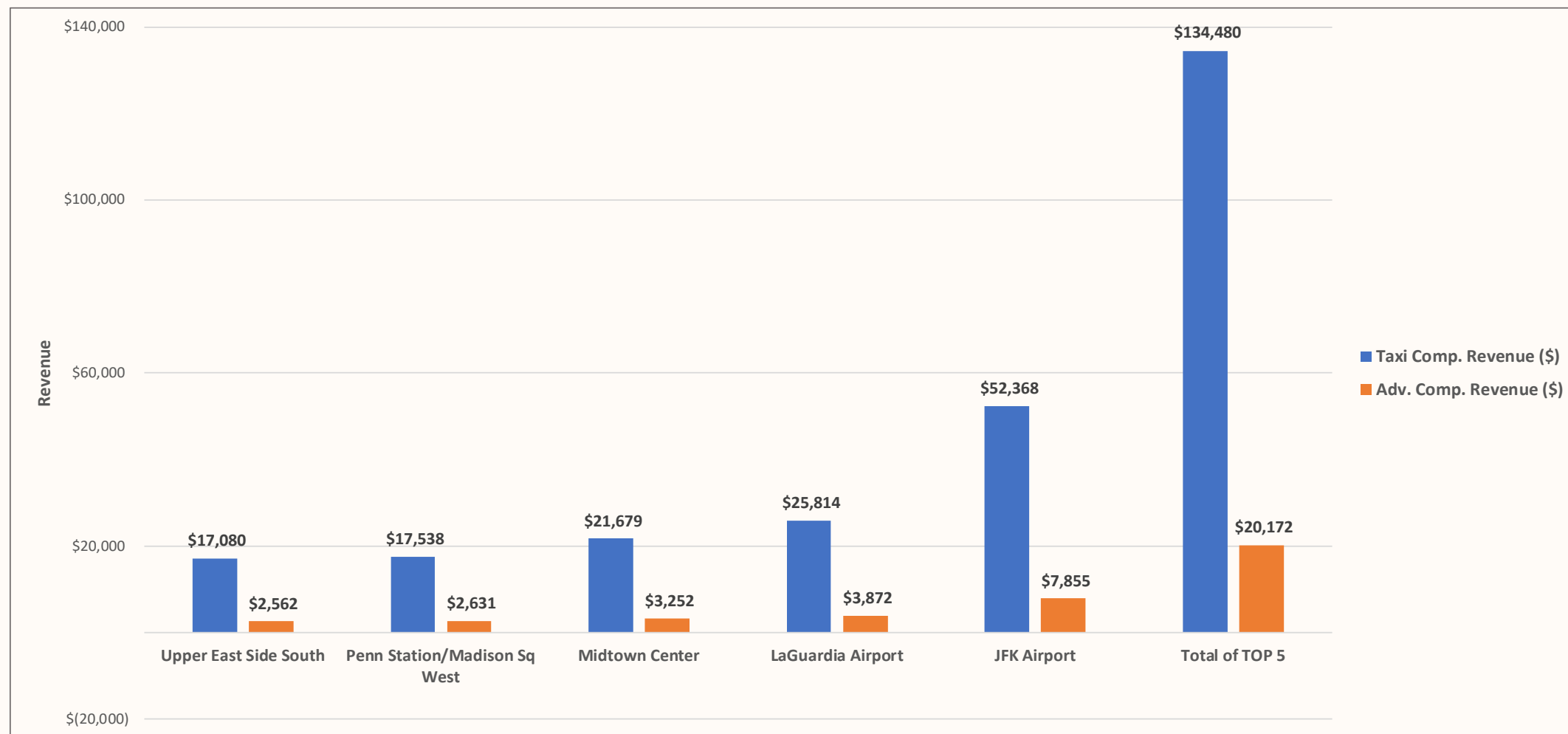
درآمد آژانس

\$۳۱۸,۲۷۵

درآمد تاکسیرانی

\$۲,۱۲۱,۸۳۸

نمودار درآمد مورد انتظار در طی سه ماه:



## سوال ۴:

شما و تعدادی از دوستانتان قصد دارید یک پلتفرم تاکسی اینترنتی برای شهر نیویورک راه اندازی کنید. یکی از مهم ترین محصولات این پلتفرم، مدل قیمت گذاری سفرهاست.

با استفاده از داده هایی که در اختیار دارید، عوامل مؤثر بر قیمت سفرهای انجام شده را بیابید و توضیح دهنده گی هر کدام را تعیین کنید. آیا الگوی قیمتی ثابتی در طول روز، هفته یا سال وجود دارد؟

بررسی کرایه سفر براساس:

**مسافت، تعداد مسافر، نوع تاکسی، شیوه تعیین نرخ، زمان (ساعت، روز، ماه، سال)**

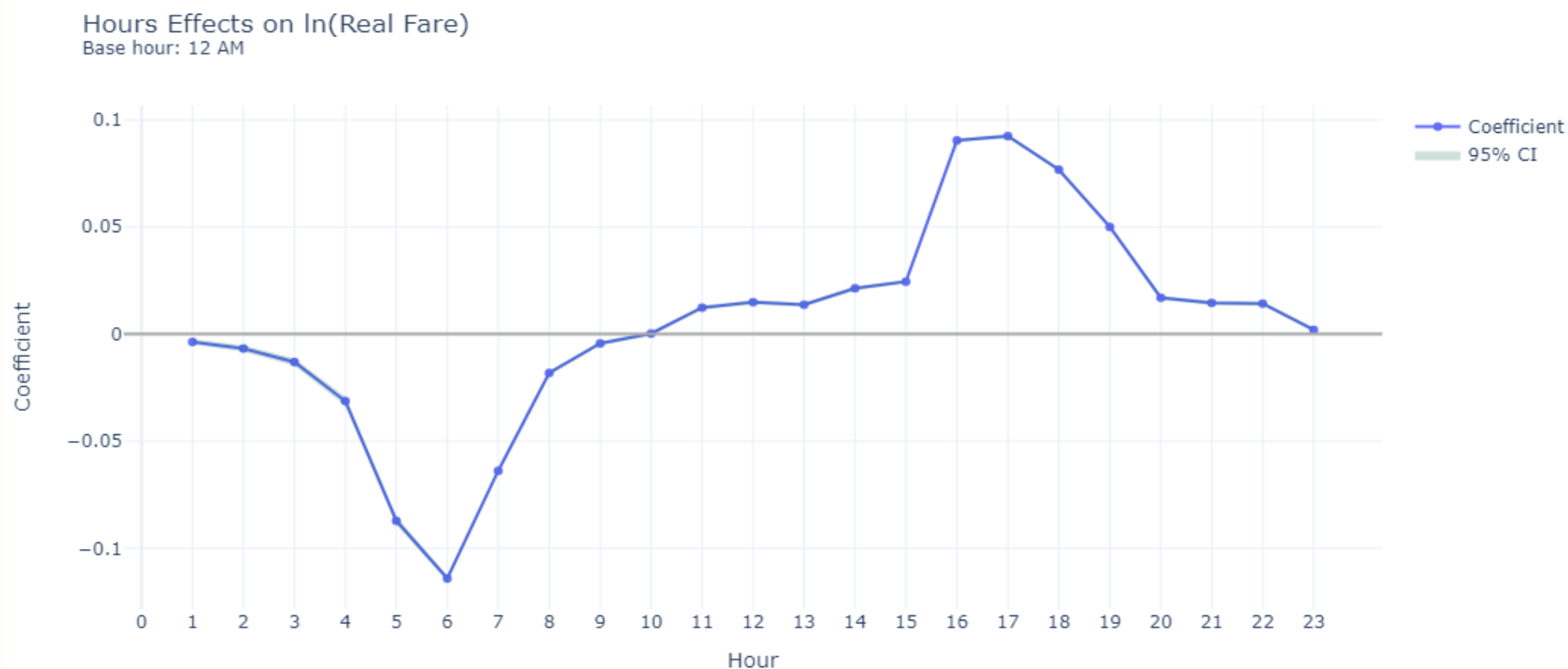
- نمونه تصادفی ۱۰٪
- محاسبه کرایه‌ی حقیقی بر اساس تورم آمریکا
- هر یک مایل مسافت موجب افزایش ۱۳٪ کرایه می‌شود
- کرایه تاکسی‌های زرد ۱۷٪ درصد از تاکسی‌های سبز بیشتر است



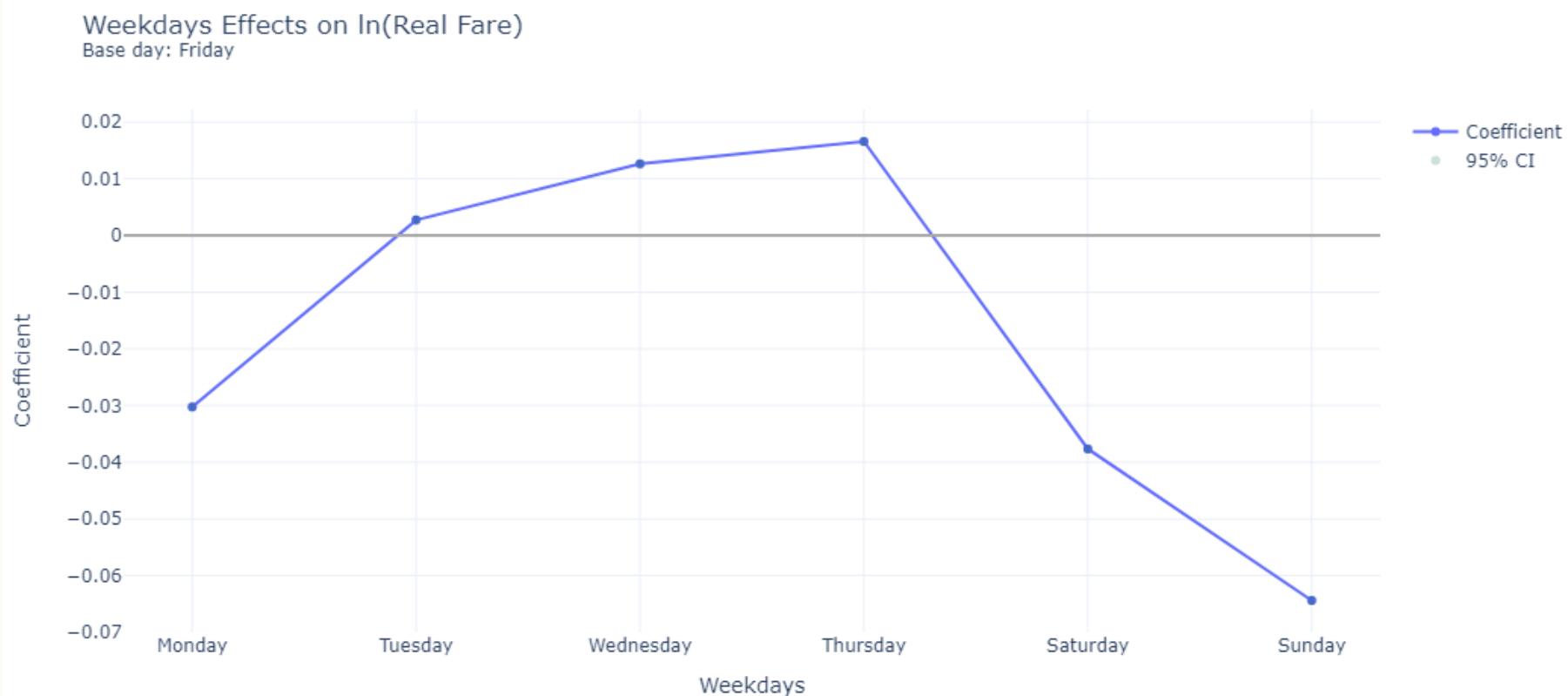
در حالی که همه روش‌های نرخ‌گذاری نسبت به روش استاندارد کرایه‌ی پایین‌تری دارند، کرایه‌ی سفرهایی که با روش مذاکره‌ای تعیین شده‌اند، گران‌تر هستند.



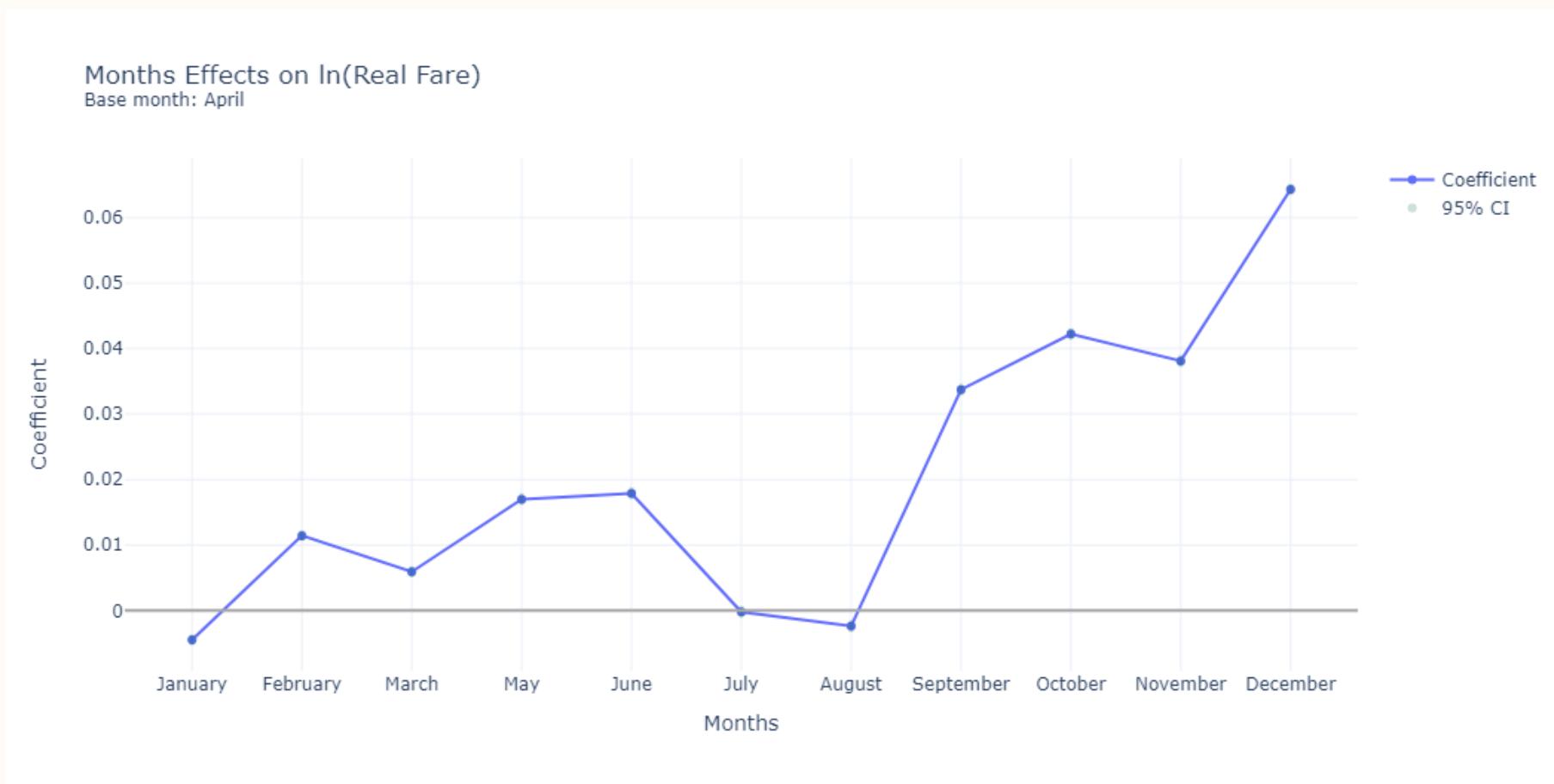
با رفتن از ساعت ۱۲ بامداد به سوی ۶ صبح کرایه سفرها به طور میانگین با افت‌های فزاینده نسبت به ساعت پایه همراه است. اما با گذر از ساعت ۶ صبح تا ساعت ۱۰ صبح این افت کم‌رنگ‌تر می‌شود.



از روز دوشنبه تا پنجشنبه کرایه‌ها روندی فزاینده را تجربه می‌کنند؛ اما با رسیدن به روزهای شنبه و یکشنبه، به عنوان روزهای تعطیل هفته، بهای حقیقی سفرها افت شدیدی را تجربه می‌کند.



به طور کلی کرایه سفرها با رفتن به ماه‌های پایانی سال افزایش می‌یابد. با این حال با عبور از ماه اوت به ماه سپتامبر کرایه‌ها رشد بالایی را تجربه می‌کنند.



## سوال ۵:

در مرحله‌ی ابتدایی راه‌اندازی این پلتفرم، لازم است محصول اولیه‌ی خود را به صورت آزمایشی برای تعدادی راننده عرضه کنید. برای این کار با تعدادی راننده توافق کرده‌اید که اپلیکیشن را روی گوشی خود نصب کنند و در محل‌های مشخصی استقرار یابند.

**الف-** محل‌هایی را برای استقرار اولیه‌ی رانندگان پیشنهاد دهید تا تیم بازاریابی بتواند در آن محل‌ها اپلیکیشن مسافر را به صورت حضوری تبلیغ و در محل برای مشتری‌ها نصب کند.

**ب-** مسافرانی که در مکان‌های پیشنهادی درخواست تاکسی دارند، عمدتاً به چه نقاطی از شهر می‌روند؟ مسیرهای پرتردد را پیدا کنید و بررسی کنید آیا این مسیرها در طول روز تغییر می‌کنند یا خیر.

- تارگت: فعالیت در مبدهایی با بیشترین حجم بازار
- تمام مدل‌های پرداختی
- بازه زمانی ۱۲ ماه اخیر
- متریک:  
(شمار سفرها × میانۀ مبلغ سفرها)
- مکان‌های منتخب (۳۰٪ حجم بازار)

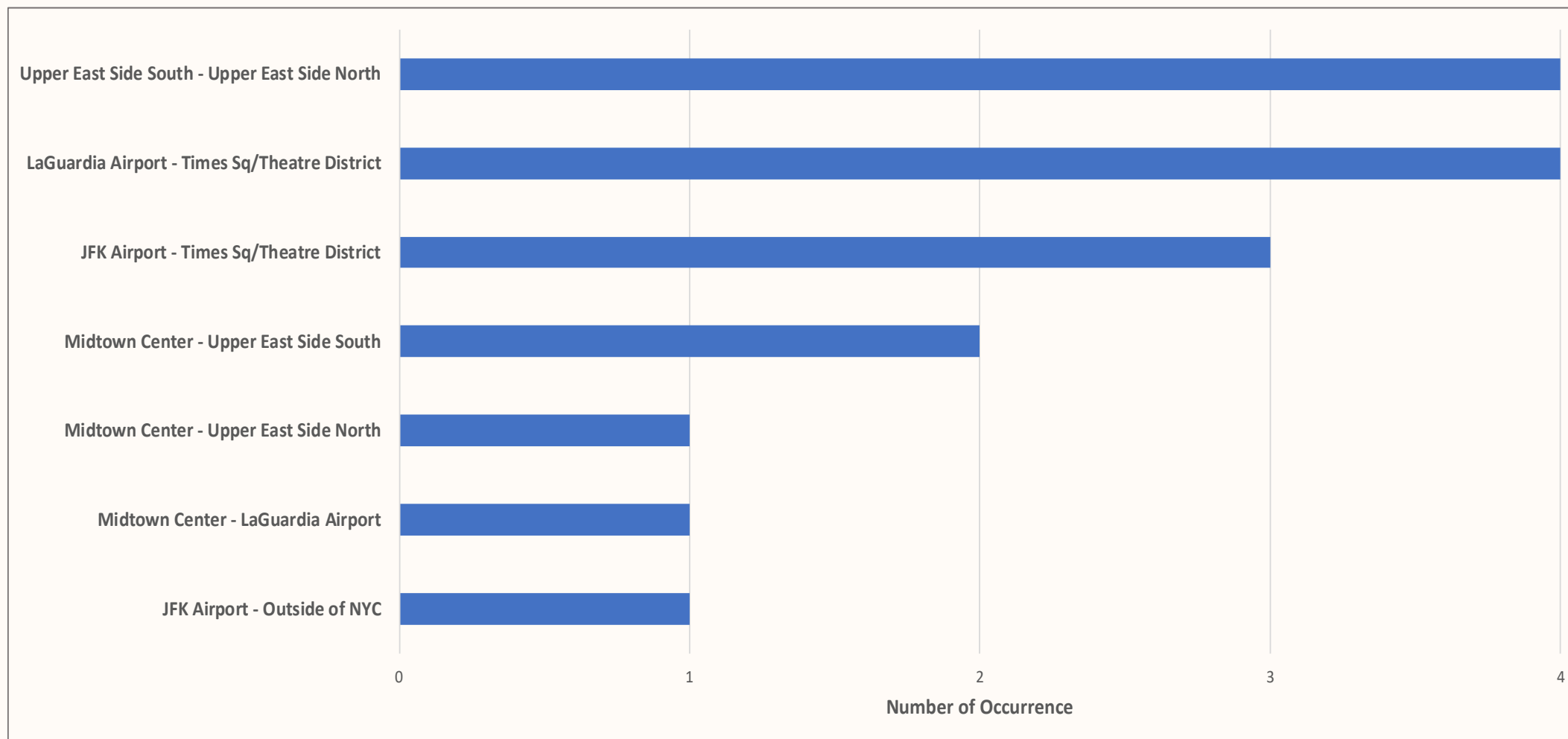
JFK Airport

Midtown Center

LaGuardia Airport

Upper East Side South

در بازه‌های زمانی ۶ ساعته تعداد تکرار مسیرهای پرتردد به صورت زیر است:



## سوال ۶:

یکی از مزیت‌های اصلی پلتفرم شما این است که می‌تواند قیمت‌گذاری غیرمنصفانه را اصلاح کند، یعنی برای مسیرهای مشابه در زمان مشابه تا حد امکان قیمت‌های مشابهی پیشنهاد دهد.

از شما خواسته شده است بزرگی این مشکل را، یعنی قیمت‌های متفاوت در مسیرهای مشابه، محاسبه کنید. آیا در طول زمان ابعاد این مشکل کم شده است؟



مسیرهای مشابه:

- مسیرهایی با بازه‌های مسافتی و بازه‌های سرعتی مشابه

زمان مشابه:

- بازه‌های ساعتی مشابه در طول شبانه روز ( تقسیم ۲۴ ساعت به ۶ دسته ۴ ساعته )

- متریک:

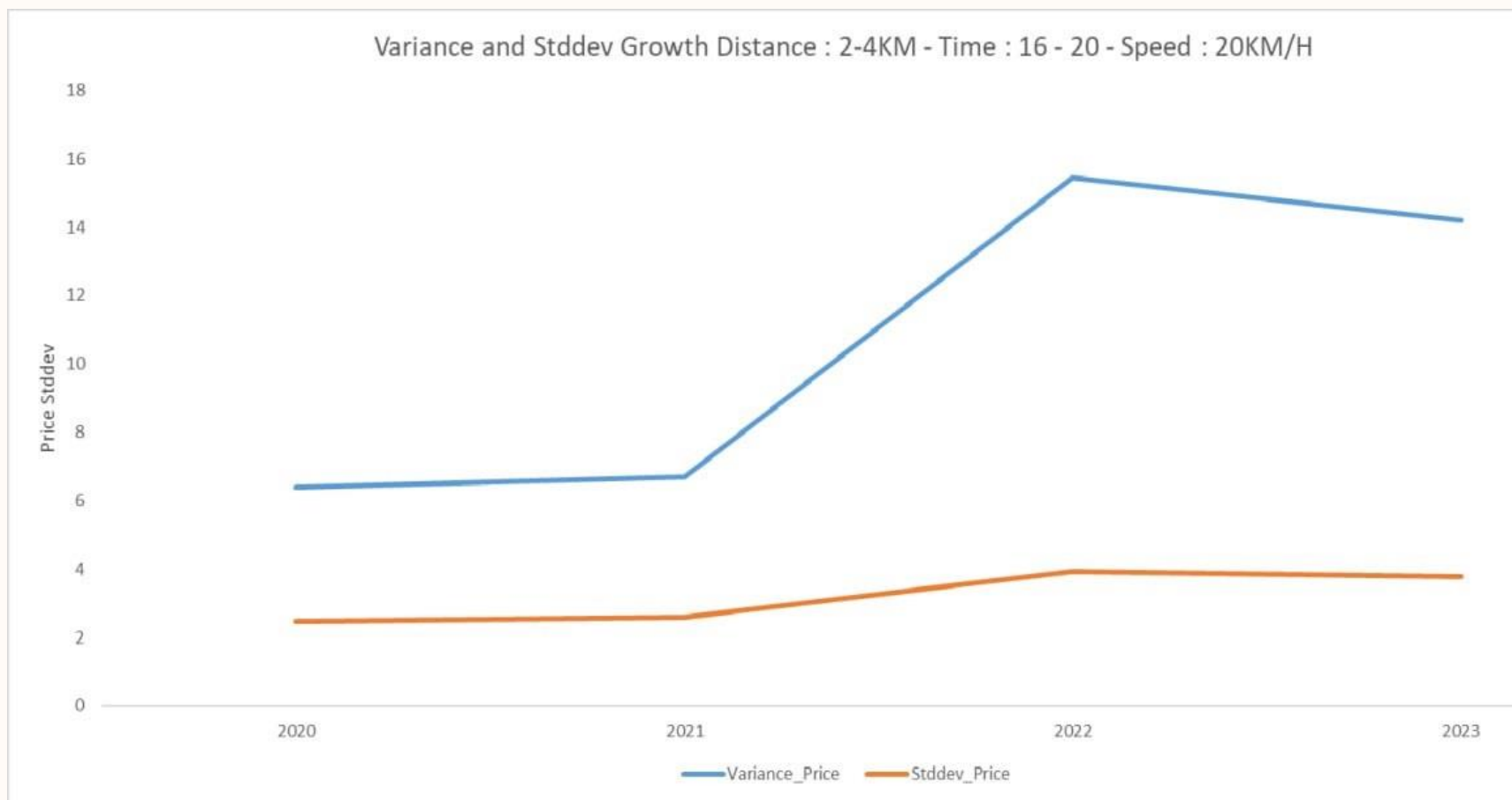
### انحراف معیار قیمت‌ها

- بازه‌های ۲ کیلومتری مسافت

- بازه‌های ۱۰ km/h سرعت

- همه‌ی کرایه‌ها به ارزش سال ۲۰۲۴ حقیقی شده

نمودار روند افزایش / کاهش و پراکندگی کرایه در ۴ سال اخیر برای یک مسیر-زمان نمونه:



مقایسه انحراف از معیار مسیر-زمان‌های مشابه در سال ۲۰۲۳ و ۲۰۲۴:

2023



2024



### نتیجه:

- مشخص است که در مسیرهای مشابه تفاوت قیمتی وجود دارد.
- با توجه به روندها می‌توان گفت که این تفاوت قیمت در یک سال اخیر کاهش یافته است.

## سوال ۷:

یکی دیگر از ایده‌های پلتفرم تاکسی آنلاین، فراهم کردن خودروهای ون برای حمل تعداد زیادی مسافر است.

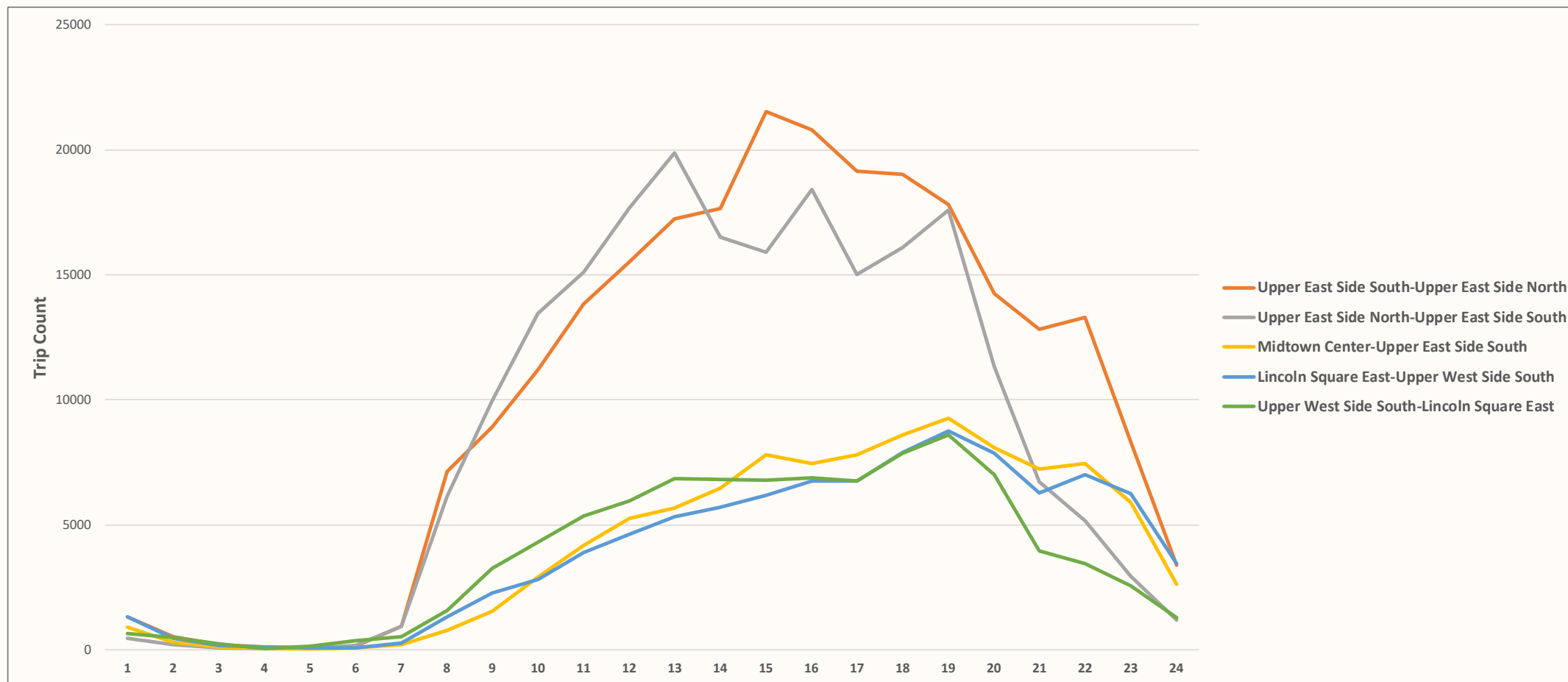
برای آزمایش این ایده، از شما خواسته شده است چند مبدأ پیشنهاد دهید و میزان تخفیفی را که می‌توان برای سفرهای ون نسبت به تاکسی ارائه داد به‌دست آورید.

- مسیرهای پرتردد
- متریک مد نظر:

### (تعداد سفر)

- توزیع مناسب سفر در هر ساعت از شبانه روز
- بازه زمانی ۱۲ ماه اخیر

## نمودار تعداد سفر در مسیرهای پرتردد در طول شبانه روز:



- تعداد مسافران هر ون ۱۰ نفر
- $X$  = مصرف سوخت تاکسی به ازای هر مایل معادل ۰.۰۳ گالن
- $Y$  = مصرف سوخت ون به ازای هر مایل معادل ۰.۰۵ گالن
- $P$  = میانگین هزینه هر گالن بنزین در سه ماه نخست ۲۰۲۴ معادل \$۳.۱۵
- حداکثر تخفیف به هر مسافر ون در ازای هر مایل مسافت:

$$\left[ x - \left( \frac{y}{10} \right) \right] \times p$$

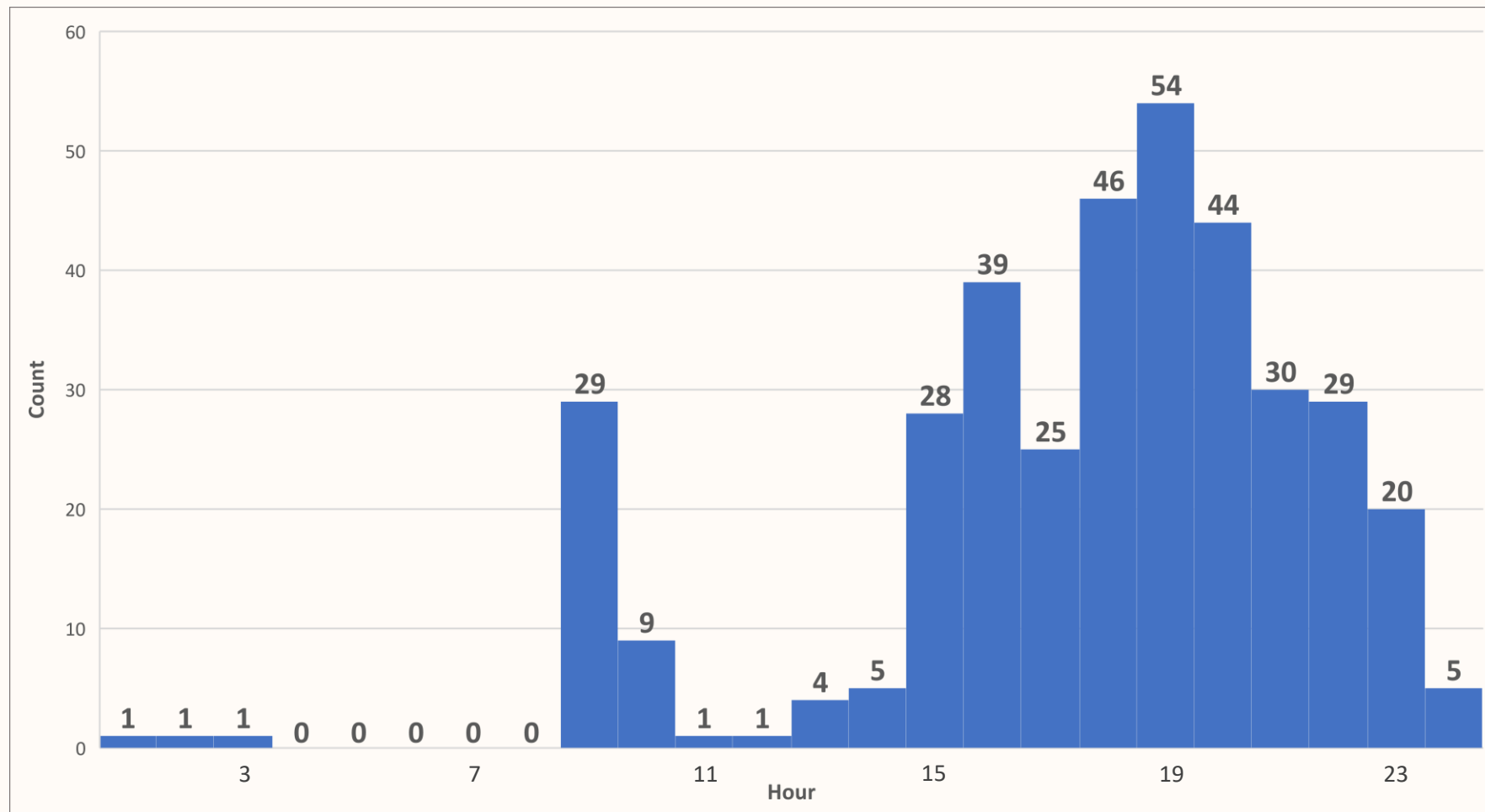


## سوال ۸:

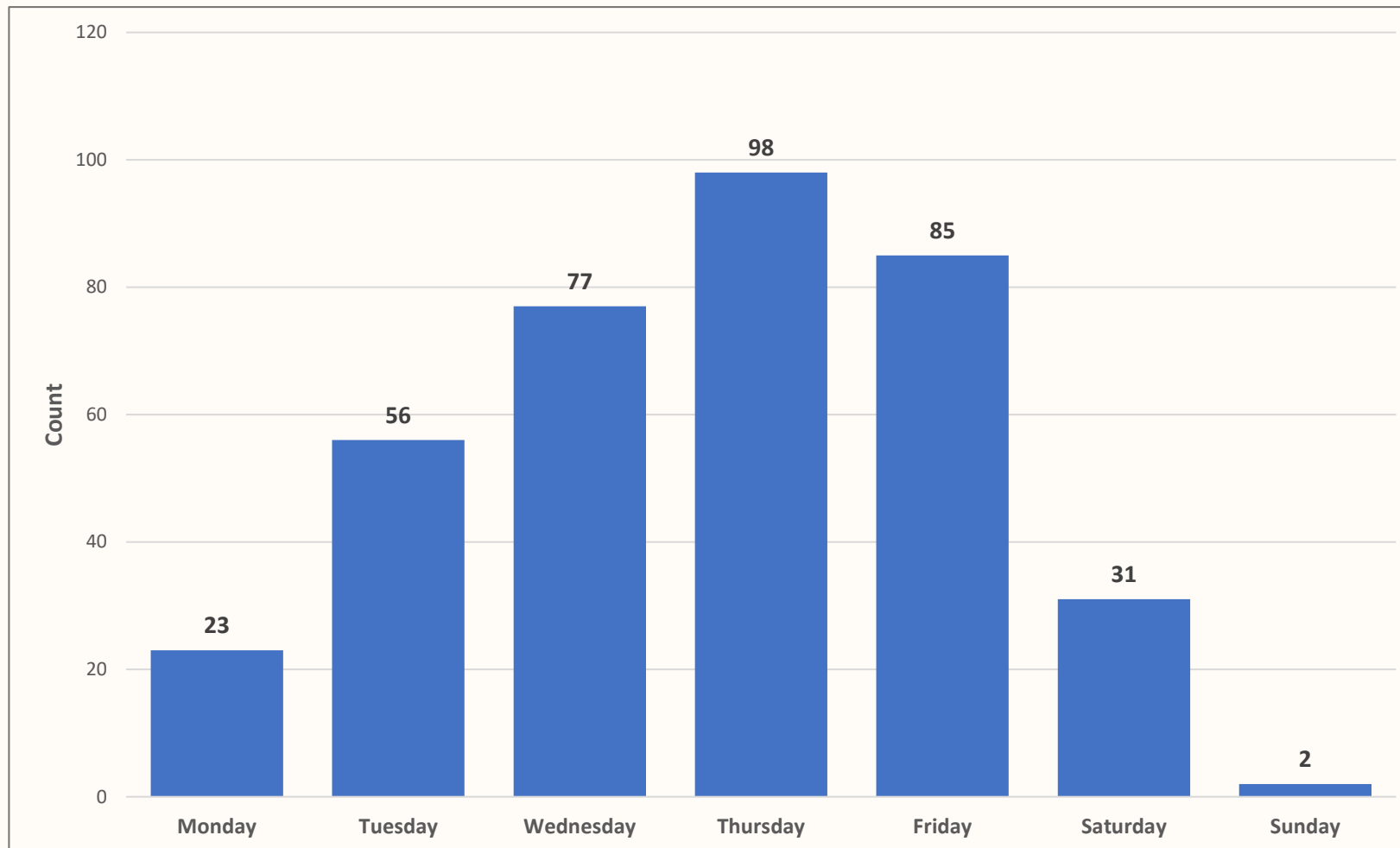
یکی از روش‌های درآمدزایی بالقوه در تاکسی‌های آنلاین توانایی پیش‌بینی آینده‌ی نزدیک و آماده شدن برای اقتضائات آن است.

در بازه‌ی زمانی مورد بررسی، ۱ درصد بالای روز-ساعت‌هایی با بیشترین تعداد سفر شروع شده را پیدا کنید. آیا می‌توانید دلیل/دلایل خاصی برای بالا بودن تعداد سفرها در این موارد بیابید؟ استفاده از داده‌های خارجی در پاسخ به این سؤال لازم است.

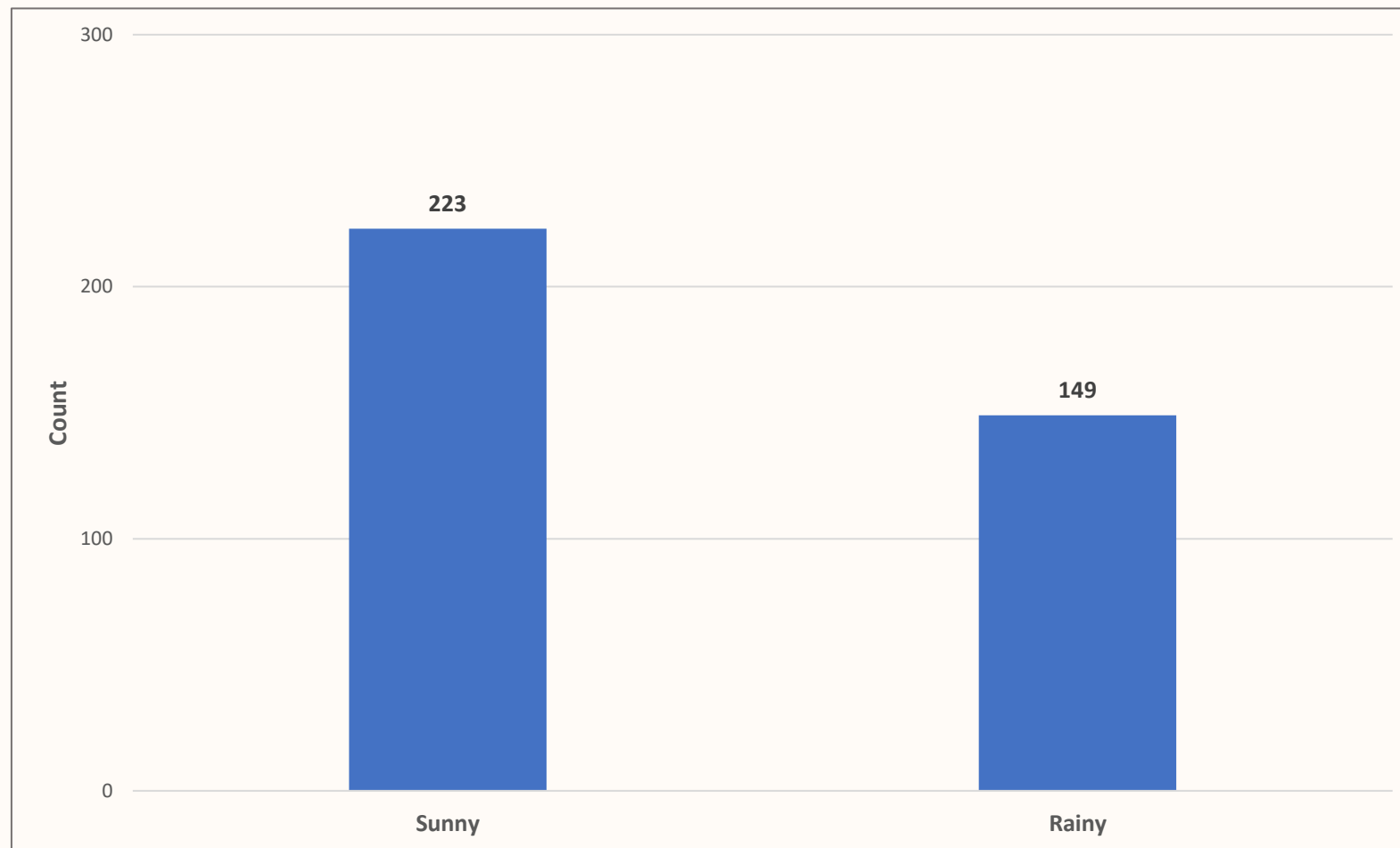
## توزیع تعداد روز-ساعت‌ها در طول شبانه روز:



توزیع تعداد روز-ساعت‌های پر تردد در طول هفته:



توزیع تعداد روز-ساعت‌های پرتردد در آب و هوای مختلف:



### نتیجه:

- در ساعات ترافیک، روزهای کاری و روزهای بارانی تقاضا برای تاکسی بیشتر است.
- عرضه بیشتر در این ساعات می‌تواند موجب درآمدزایی بیشتر شود.
- می‌توان برای تشویق راننده‌ها برای کار در این ساعات، مشوق‌هایی در نظر گرفت.



ممنون از وقتی که گذاشتی!