فهم مسئله:

1. انگیزه اصلی چنین پروژهای چیست؟

هدف از انجام چنین پروژه هایی جمع آوری اطلاعات از عملکرد گذشته ودستیابی به آمار خاص و یا تحلیل و بررسی آنها برای تصمیم گری های بهینه برای رفع مشکل موجود و یا بررسی موانع پیشبرد فعالیت و همچنین ایجاد یک دید نسبت به رفتار های کنونی می باشد.

1. خروجی چنین پروژهای برای چه مواردی ممکن است کاربرد داشته باشد؟

* پزشکی
* بازار دیجیتال
* کشاورزی
* ورزش
* سینما و تلویزیون
* آموزش
* حوزه مالی
* حمل و نقل
* محیط زیست
* ساخت و تولید
* عملکرد شرکتها و فروشگاهها برای جذب مشتری

1. چه کسانی ممکن است به نتایج این پروژه علاقمند باشند؟ چرا؟

* تمامی افراد جامعه به نوعی می تونند علاقمند باشند.
* هر کدام در حیطه فعالیت و علایق و کار خود و حتی بهبود کیفیت زندگی و تربیت و آموزش کودکان و رفع نیازهای روزمرگی زندگی خود .

درک داده:

1. دادهها از کجا بدست آمده اند و چگونه جمع آوری شده اند؟

دادها در یک فایل csv جمع آوری شده بودن

1. هر یک از متغیرها چه چیزی را اندازه گیری میکنند؟

میانگین موارد موثر در قیمت خانه

1. آیا ابهامی در تعاریف داده ها وجود دارد؟

خیر

1. آیا ممکن است در اندازه گیری متغیرها و یا ثبت داده ها خطایی وجود داشته باشد؟

به همین دلیل وجود داده پرت، (Outlier Data) یا دورافتاده، که ممکن است براثر خطا اندازه‌ گیری بوجود آمده یا واقعا از اعضای استثنایی جامعه باشند، که همین شاخص‌ها ی اندازه گیری را به شدت تحت تاثیر قرار می‌دهند.

1. چه متغیرهای دیگری اگر وجود داشتند، میتوانست به حل مسئله کمک کند؟

* محل قرار گرفتن خانه در شهر
* تجهیزات استفاده شده برای ساخت و یا اسکلت خانه
* تعداد طبقات موجود در خانه
* وجود تجهیزات سرمایشیو گرمایش

1. متغیرهای موجود از کدام نوعند(رستهای\_عددی)؟

فقط آدرس از نوع غیر عددی بود و بقیه عددی بودن

1. خلاصه آماری متغیرهای موجود چیست؟

آماده سازی داده:

1. آیا نیاز به درآمیختن داده ها است؟ اقدامات و نتایج گزارش شود.

خیر

1. آیا نیاز به پاکسازی داده است؟ اقدامات و نتایج گزارش شود.

فقط آدرس که از نوع object بود

1. آیا نیاز به تبدیل داده است؟ اقدامات و نتایج گزارش شود.

خیر

1. آیا نیاز به کاهش داده است؟ اقدامات و نتایج گزارش شود.

خیر

مدلسازی:

1. روی داده های آموزش ،الگوریتم رگرسیون ساخته شود .اقدامات و نتایج گزارش شود.

ارزیابی:

1. مدلهای ارائه شده، روی دادههای آزمایش با استفاده از شاخصهای ارزیابی رگرسیون خطی 1 در یادگیری ماشین ارزیابی شوند .اقدامات و نتایج گزارش شود.
2. چه پیشنهاداتی دارید تا نتایج در محیط واقعی، آزمایش گردد؟

استقرار:

حال اگر بخواهید چنین الگوریتمی را در مقیاس صنعتی توسعه دهید، به این فکر کنید با چه چالش هایی مواجه خواهید شد و برای آن چه راهکارهایی دارید. موارد زیر را گزارش کنید:

1. چالشهای توسعه الگوریتم را بررسی کنید.
2. چه راهکارهایی برای حل آن ها دارید؟
3. چه ملزوماتی برای ارائه آن راهکارها نیاز دارید؟

نتیجه گیری:

1. انجام این پروژه چه یادگیری برای شما داشت؟

آشنایی با کتابخانه های numpy, pandas, matplotlib , و رسم نمودار های مختلف و مشاهده و یادگیری کاربد عملی این کتابخانه ها

1. با چه چالشهایی مواجه شدید؟ چگونه آنها را حل کردید؟

به نظر من امکان دسته بندی داده ها وجود نداشت .