

پاسخ ها چالش تحلیل داده:

Section 1:

- ۱- داده ها با من صحبت از الگوهای مستتر داخل رفتار و نشون میده علاقه من بیشتر برای یافتن اتفاق ها از طریق این الگو ها هستم. و علاقه که در این قسمت صنعت نرم افزار است بیشتر برای من هیجان انگیز است تا توسعه و معماری های مختلف نرم افزاری. بیشتر می توان به کمک ریاضیات که علاقه اولیه خودم بوده بتوانم الگوها و اتفاق هایی را از طریق داده ها بدست بیاورم.
- ۲- اتفاق ها که داخل یک مجموعه زندگی عموم بطور مثال شهر شاید بدست نیامده. ارزشمندی داده ها در هر شاخه ای میتونه کمک کنه از کم کردن ترافیک پیشگیری از بیماری ها بهبود کیفیت زندگی
- ۳- مهمترین فرق تحلیل گر داده با مهندس داده در این است که مهندس داده بیشتر در جمع اوری داده های مورد نیاز یک تحلیل گر می تواند کمک کند که در کنار اینکه داده های جمع اوری شده را می تواند برای یک دانشمند داده نییز جمع اوری کند . دانشمند داده با تکیه اصلی بر دانش ریاضیات احتمالات میتواند تا حدودی آیش بینی و تشخیص دهد.
- ۴- 25 میلیون رکورد مهارت ها تسلط به زبان برنامه نویسی و sql و دانش ریاضیات
- ۵- داده ها نباید تکراری باشند. باید بتوان به منبع اصلی ایجاد داده رسید. داده ها باید یک محدوده زمانی داشته باشند یعنی نباید داده های مورد بررسی تایم های ان تفاوت های زیادی باهم داشته باشند.
- اماده کردن داده ها اعم از حذف تکراری ها و خطا دار بودن داده ها. داده هایی که در نرم نیست نیز باید جدا شوند.

secction2

تمامی پاسخ ها سوال ها با محاسبات در دیتابیس انتخابی postgres انجام شد
سوال اول میزان محبوبیت در فایل ضمیمه شده پاسخ داده شده است

سوال دوم نیز در فایل ضمیمه شده پاسخ داده شده است.

سوال سوم : از جدول friendship تعداد ارتباط هر کاربر را شمارش شد. بیشتر بصورت اینکه با تابع count در sql با پارانشن بندی بر اساس هر کاربر محاسبه شد هر کاربر با چه تعداد کاربر دیگری در ارتباط است. ۵ نفر اول پرتعامل به ترت

sara	7
zahra	6
siamak	5
narges	5
mahsa	5

و همچنین میزان محبوبیت ۵ نفر اول:

simak	50
raha	49
negin	48
aida	48
shahab	47