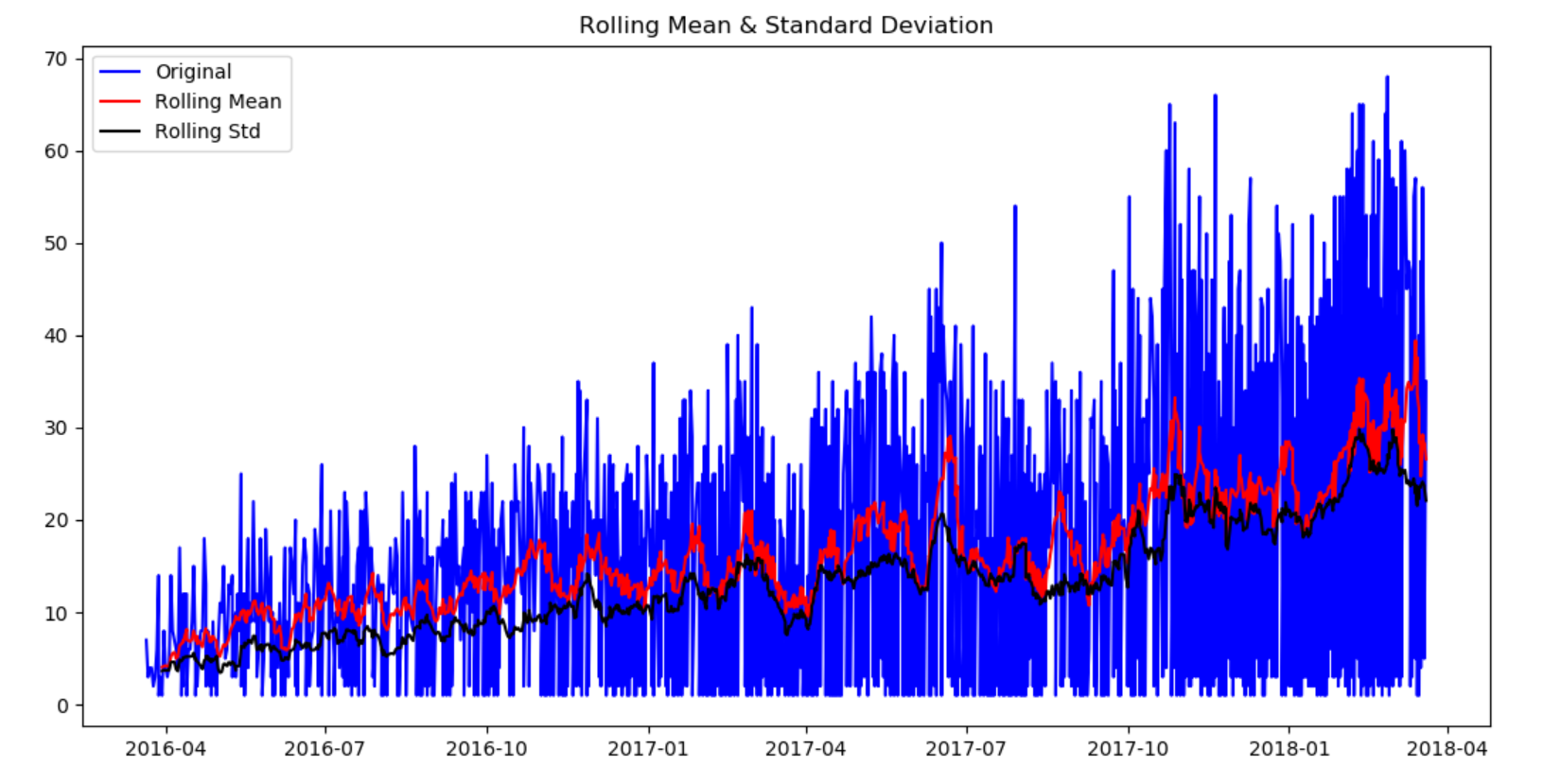
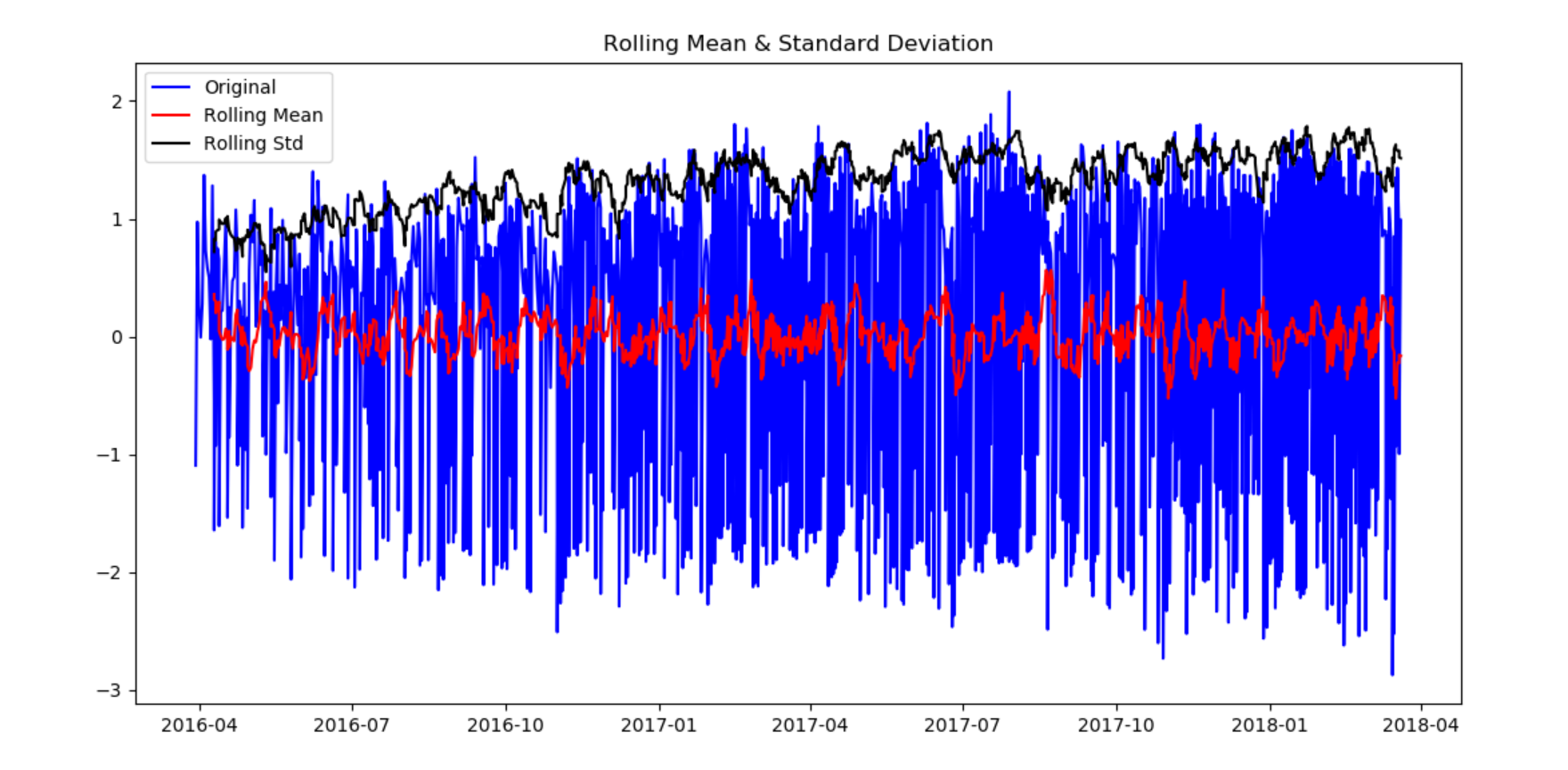
عملکرد گروه ما به طور کلی شامل چندین فاز مختلف بود :

مرحله اول درک مسیله بود . در این مرحله ، با بررسی داده های آموزش و همین طور جستجو در منابع مختلف تلاش کردیم تا مثال های مشابه و ایده ها و راه حل های موجود را بیابیم. در مرحله دوم پس از بررسی منابع یافته شده تلاش کردیم که داده های مسیله را به حالت استاندارد در بیاوریم و برای این کار در یک گام داده ها را به صورت یک جدول جدید با 4 ستون با نام های مبدا (FROM) و مقصر (TO ) و تاریخ پرواز (Log\_date) و تعداد دفعات پرواز (count) در آوردیم . و در گام بعد تاریخ های شمسی را به تاریخ های میلادی (برای استاندارد سازی داده ها و توانایی بهره جویی از قابلیت index و امکانات موجود در کتابخانه ها) تبدیل کردیم. پس از آماده سازی و انجام preprocess های گفته شده تیم ما تصمیم گرفت برای هر سفر یک مدل جداگانه بسازد و برای تخمین از همان مدل استفاده کند. پس ، یک تابع ایجاد کردیم که به ازای گرفتن ورودی مبدا و مقصد ، اطلاعات موجود به تنها این خط پروازی را میدهد. برای آزمایش و مدل سازی ، از یکی از خط های پروازی که مشتری نسبتا زیادی داشتند (خط پروازی از 75 به 70) استفاده کردیم و سپس برای تمامی خطوط ایجاد کردیم.

در ادامه برای بررسی وضعیت کلی داده از دو معیار rolling mean که در واقع میانگین در 12 ماه گذشته و rolling standard که همان deviation در 12 ماه گذشته است استفاده میکنیم. حاصل خروجی کار این مرحله آماده یازی داده ها به صورت زیر بود :

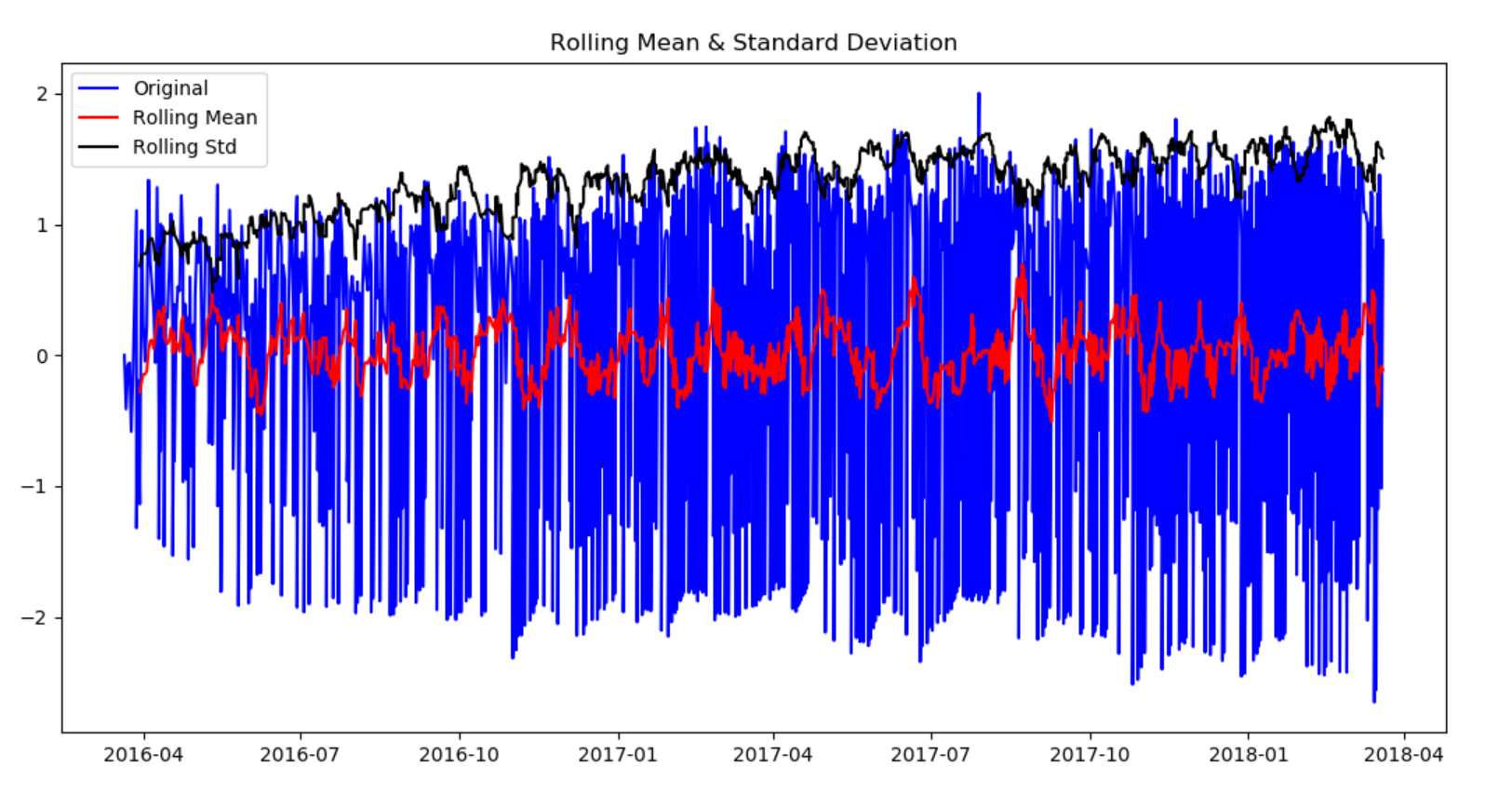
در فاز بعد تلاش شد تا داده ها را به صورت سری زمانی ثابت دربیاوریم. با بررسی و تحقیقات به عمل آمده به این نتیجه رسیدیم که دو دلیل عمده وجود دارد ، اولی تغییرا ت در روند کلی داده ها که در این جا نیز وجود دارد و همان طور که دیده میشود متوسط فروش کلی داده ها در این دوسال افزایش داشته است. و دلیل دوم وجود تغییرات در فریم های زمانی خاص که در این جا همان طور که میبینیم در بعضی تاریخ ها فروش بسیار زیاد شده و دوباره به نرخ سابق برگشته است.

برای از بین بردن مشکل اول میتوان از روش های مختلف همانند Aggregation ، Polynomial Fitting و همین طور Smoothing استفاده کرد . که در این جا با بررسی های انجام شده ، روش Smoothing انتخاب شد. در این روش اول از میانگین 12 ماه قبل استفاده میکنیم و این مقدار را از داده های اصلی کم میکنیم. پس از انجام این عملیات داده ها به این صورت درمیایند :

\*\*قابل توجه است که در این جا ابتدا از از تبدیل لگاریتم برای transformation داده ها استفاده کرده ایم

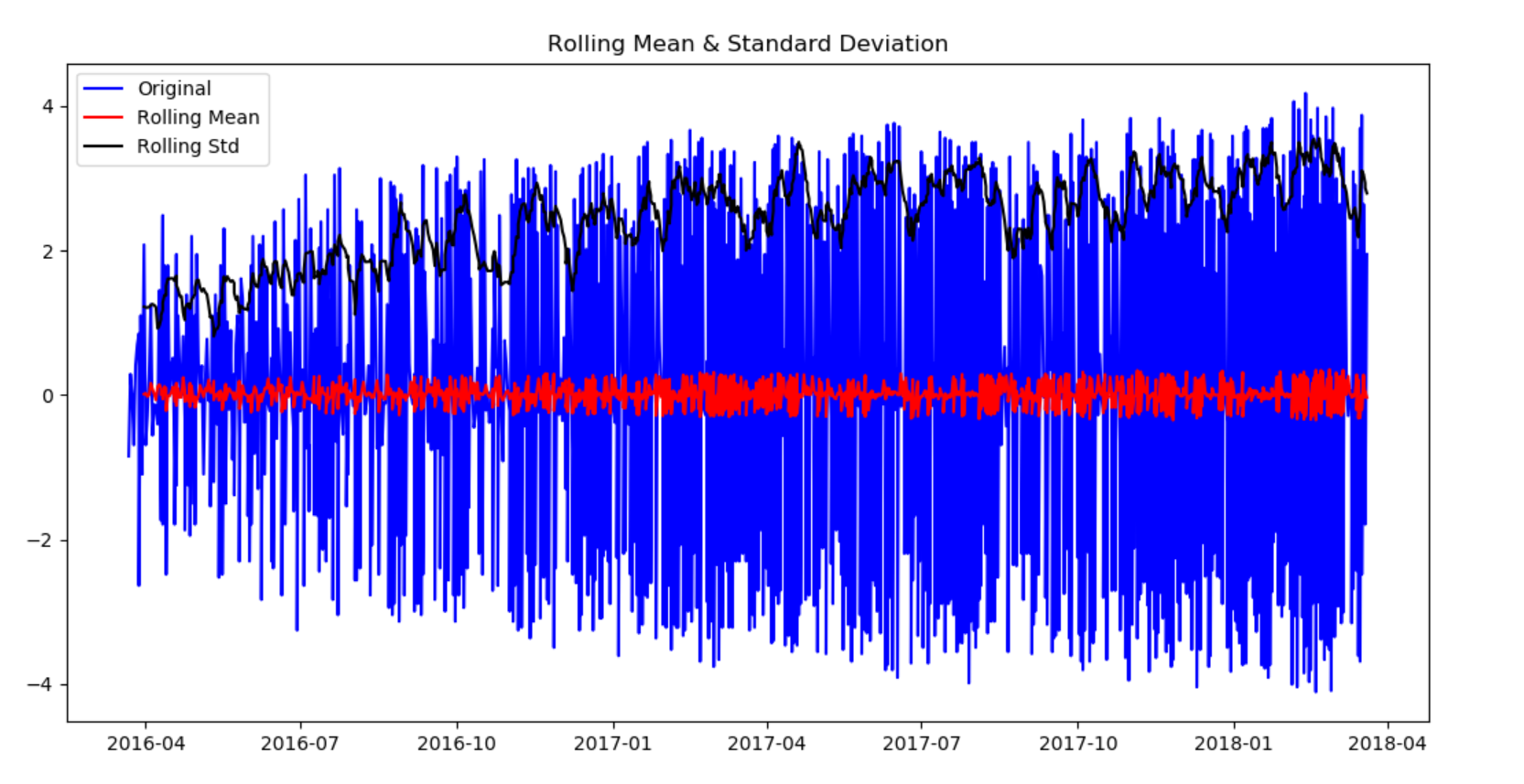
همان طور که میبینیم، نرخ رشد کلی داده ها تا حدی از بین رفته است .

اما برای دقیق تر شدن این کار میتوان از میانگین وزن دار استفاده کرد و برای داده های جدید تر ارزش بیشتری قایل شد. برای این کار از exponentially weighted moving average استفاده میکنیم و در این صورت میبینیم که این داده ها تا حد بسیار خوبی به سری های زمانی منطبق شده اند و test statistic ما از 1% داده ها نیز کمتر شده که این به این مفهوم است که با اطمینان 99% میتوانیم بگوییم که این سری یک سری زمانی است. این نمودار به صورت زیر خواهد بود :

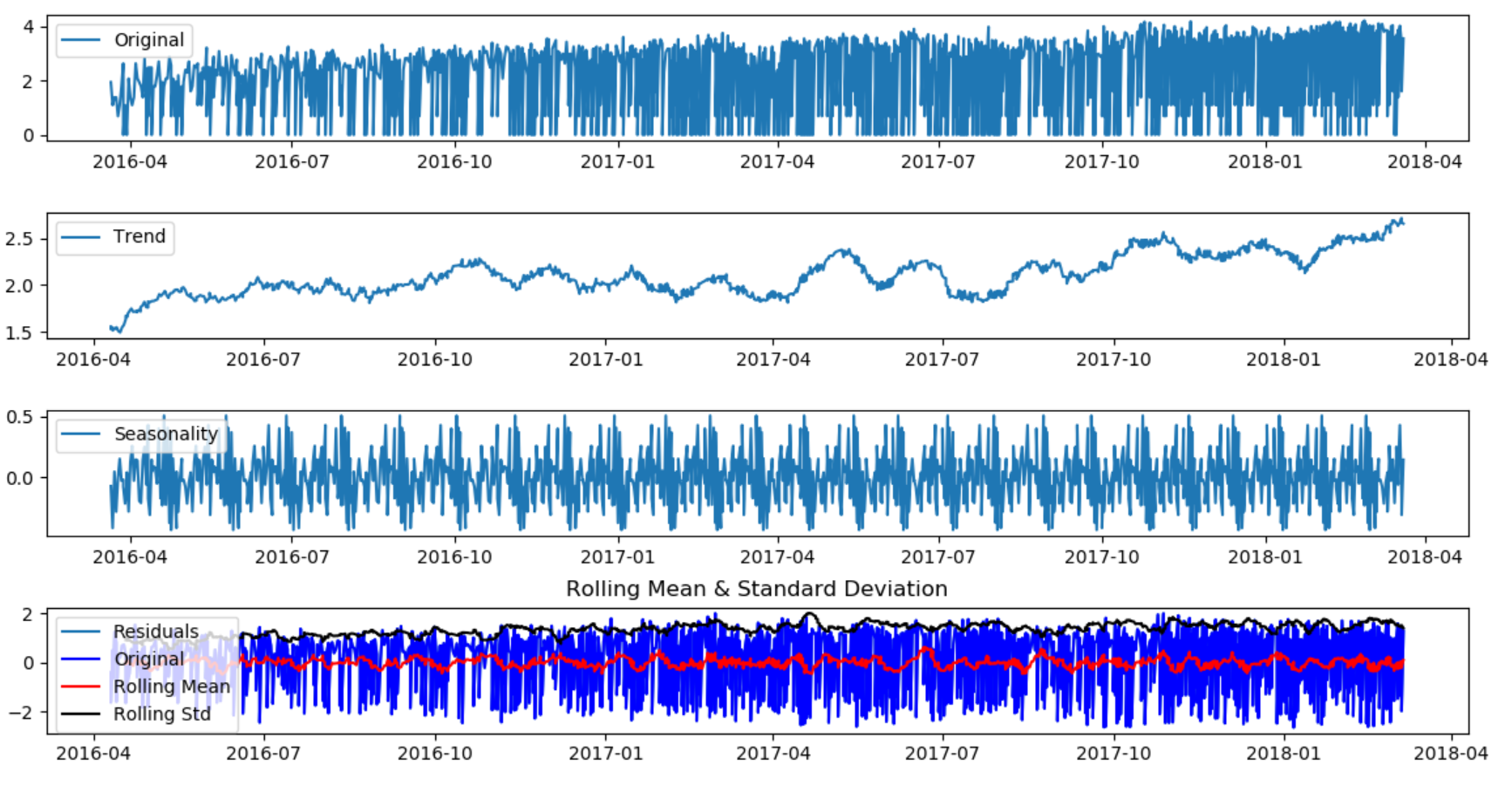


اما این راه حل هر چند خوب نمیتواند اثری در مشکل دوم گفته شده که همان تغییرات در دوره های زمانی خاص یا seasonality است داشته باشد . برای حل این مشکل باید به سراغ راه حل هایی برویم که هر دو مشکل را حل کنند. در این جا از دو روش میتوان استفاده کرد.

راه اول این است که از تفاضل داده ها در بازه های زمانی مشخص استفاده کنیم و راه حل دوم به نام decomposition این است که مدلی برای trend ها و یا رشد کلی و مدلی برای seasonality بسازیم و این دو را از داده های اصلی کم کنیم. که حاصل راه حل اول به صورت زیر خواهد بود :



که کمک چندانی به ما نکرد و حتی نتوانست تا 90% تطابق را ایجاد کند.

و راه دوم حاصلی به صورت زیر داشت:

که test stationary نسبت به وضعیت روش قبل بهبود داشت اما باز هم نسبن به وضعیت روش exponentially weighted قطعیت کمتری داشت . پس ترتیب اثرگذاری روش ها به این صورت شد که اثرگذارترین روش exponentially weighted moving average سپس روش decomposition سپس روش moving average با ضرایب ثابت و در نهایت روش differencing . پس در مراحل بعد از این روش استفاده میکنیم.

# اقا من چیزایی که تا اینجا فهمیدمو نوشتم و سعی کردم که یه جوری بنویسم که انگار واقعا خودمون انجام دادیم و فهمیدیم . فقط چند تا چیز که موقع نوشتن فهمیدم .

1. ما باید این مدل ها رو برای همه داده درست کنیم و بذاریم توی یه آرایه ای و برای تست با توجه به مبدا مقصد از مدلش استفاده کنیم
2. نباید برای ساخت مدل از داده های همین جوری استفاده میکردیم و باید از اون تغییراتی که توی plotter انجام میدادیم استفاده کنیم که من چک کردم بهترین نتیجه برای ما روی روش exponentially بود و باید از اون برای ساخت مدل اینا استفاده کنیم
3. فهمیدم هم که این test stationary چی بود ، توضیحشو تا حدودی بالا نوشتم ، اگه نفهمیدین بگین دوباره بگم.