

دانشگاه صنعتي امیرکبیر  
(پلی تکنیک تهران)

دانشكده مهندسی کامپیوتر

گزارش پروژه

پایش رویداد‌های فضای مجازی

نگارش

بردیا اردکانیان

محمدرضا حسن‌زاده

موژان میرجلیلی

فرحان کیهان

استاد راهنما

دکتر میثم نظریانی

استادمشاور

دکتر میثم نظریانی

خرداد ۱۴۰۱



# سپاس‌گزاری

در اینجا بر خود لازم می‌دانیم که از راهنمای‌های استاد گرانقدر جناب اقای دکتر نظریانی کمال تشکر را داشته باشیم و همچنین از تمام کسانی که کمکان کرده‌اند، اظهار قدردانی می‌کنیم.

# چكيده

در سال‌های اخیر بورس اوراق بهادار مورد توجه افراد زیادی قرار گرفته است. علت این موضوع می‌تواند از شرایط اقتصادی تا مناسب بودن شرایط سرمایه‌گزاری در بازار اوراق بهادر باشد. به همین سبب افرادی که تجربه کمتری در زمینه بورس داشتند به دنبال افراد/افرادی گشتند که چراغ راهشان در این بازار جدید باشد.

افراد بسیاری در شبکه‌های مجازی با نام تحلیلگر بورسی پیشنهاداتی مبنی بر خرید یا فروش سهام‌ها می‌دهند. گاها افرادی پیدا می‌شوند که با سوء استفاده از «اطلاعات نهانی» و «سیگنال دهی» نسبت به سهام داران برتری پیدا می‌کنند. این سوء استفاده باعث کاهش اعتماد عمومی نسبت به بازار اوراق بهادار می‌شود.

در این پروژه به بررسی چند طرح برای بررسی و کشف تقلب در بازار اوراق بهادار می‌پردازیم. طرح مطرح شده در این پروژه روش‌های اخیر و مرسوم دنیای دیجیتال می‌باشد که در این پروژه علاوه بر الگوریتم‌های این روش‌ها برتری هر کدام را مورد بررسی قرار می‌دهیم که باعث جلب اعتماد عمومی نسب به بازار اوراق بهادار خواهد شد.

واژه‌های کلیدی:

بازار اوراق بهادار، سیگنال دهی، کشف تقلب، اطلاعات نهانی، پایش

|  |  |
| --- | --- |
| فهرست مطالب | صفحه |

[سپاس‌گزاری 1](#_Toc107952563)

[چكيده ‌أ](#_Toc107952564)

[فصل اول مقدمهمقدمه 1](#_Toc107952565)

[2-1- اهداف پروژه 2](#_Toc107952566)

[2-3-1- اهداف آرمانی 2](#_Toc107952567)

[2-3-2- هدف کلی 2](#_Toc107952568)

[2-3-3- اهداف ویژه 2](#_Toc107952569)

[2-3-4- اهداف کاربردی 2](#_Toc107952570)

[2-2- راه‌حل ایده‌آل و نیازمندی‌های مرتبط با پایش رویداد‌های فضای مجازی به منظور نظارت بر بازار سرمایه 3](#_Toc107952571)

[2-3-1- فضا‌های مجازی هدف 3](#_Toc107952572)

[2-3-2- نیازمندی‌های مرتبط با پایش رویداد‌های فضای مجازی 3](#_Toc107952573)

[2-3-3- تحلیل به‌منظور نظارت بر بازار سرمایه 5](#_Toc107952574)

[2-3- تحلیل پیام تصویری 6](#_Toc107952575)

[2-4- تحلیل پیام متنی 6](#_Toc107952576)

[فصل دوم مشخص کردن محدوده و ساده‌سازی پروژه 7](#_Toc107952577)

[تفاوت پروژه ایده‌آل و پروژه هدف 8](#_Toc107952578)

[2-1- مشخص کردن Scope و Not Scope 8](#_Toc107952579)

[2-3-4- محدوده پروژه 8](#_Toc107952580)

[2-3-5- خارج محدوده پروژه 9](#_Toc107952581)

[2-2- معرفی نیازمندی‌های شناسایی‌شده در قالب Product Backlog 9](#_Toc107952582)

[2-3-6- Product Backlog مربوط به تیم توسعه محصول 10](#_Toc107952583)

[2-3-7- تیم تحلیل‌گر 11](#_Toc107952584)

[2-3-8- تعریف مفاهیم Acceptance Criteria و Definition of Done 11](#_Toc107952585)

[2-3- تعریف مفاهیم برای هر Product Backlog Item 12](#_Toc107952586)

[2-4- برنامه‌ریزی اسپرینت‌ و جلسات اسکرام 13](#_Toc107952587)

[2-3-9- مکان جلسات 14](#_Toc107952588)

[2-3-10- مراحل اتمام هر Product Backlog 14](#_Toc107952589)

[2-3-11- اسپرینت اول پیاده‌سازی پایگاه‌داده، بات‌نت و استخراج نماد‌ها. 15](#_Toc107952590)

[2-5- عوامل پیروزی در اسپرینت اول: جلسات روزانه 16](#_Toc107952591)

[2-6- عوامل پیروزی در اسپرینت اول: کار کردن در محیط اشتراکی 16](#_Toc107952592)

[2-3-12- اسپرینت دوم پیاده‌سازی تحلیل‌گر و تحلیل متخصصین 16](#_Toc107952593)

[2-3-13- اسپرینت سوم تحلیل و خروجی نهایی 18](#_Toc107952594)

[فصل سوم چالش‌های تیم توسعه محصول 21](#_Toc107952595)

[ایده اولیه 22](#_Toc107952596)

[2-7- چالش‌های مربوط به پایگاه داده 22](#_Toc107952597)

[2-8- چالش‌های مربوط به‌صف کار 22](#_Toc107952598)

[2-9- چالش‌های مربوط به استفاده از پایتون به‌صورت خام 22](#_Toc107952599)

[2-10- چالش‌های مربوط به ربات تلگرام 23](#_Toc107952600)

[2-3-14- مهیاسازی و آماده‌سازی‌های اولیه 23](#_Toc107952601)

[2-3-15- پیاده‌سازی ربات تلگرامی با استفاده از Telethon 23](#_Toc107952602)

[2-3-16- چالش‌های پیاده‌سازی ربات تلگرامی 25](#_Toc107952603)

[2-11- چالش‌های پیاده‌سازی پایگاه داده 25](#_Toc107952604)

[2-12- چالش‌های مربوط به تحلیلگر متن 25](#_Toc107952605)

[2-3-17- پیاده‌سازی تشخیص سیگنال 26](#_Toc107952606)

[2-3-18- بهبود و افزایش کارایی الگوریتم 28](#_Toc107952607)

[2-3-19- چالش‌های پیاده‌سازی تحلیلگر پیام 28](#_Toc107952608)

[فصل چهارم چالش‌های تیم تحلیل چالش‌های مربوط به نرم‌افزار Power BI 29](#_Toc107952609)

[2-13- چالش‌های سخت‌افزاری مربوط به تیم تحلیل 30](#_Toc107952610)

[2-14- چالش‌های مربوط به ایجاد حساب 31](#_Toc107952611)

[2-15- چالش‌های مربوط به فرمت ورودی Power BI 31](#_Toc107952612)

[2-16- آموزش و یادگیری Power BI 31](#_Toc107952613)

[2-3-20- یادگیری با داده آموزشی 31](#_Toc107952614)

[2-17- امکانات Power BI 33](#_Toc107952615)

[2-18- ترکیب کردن کوئری‌ها 33](#_Toc107952616)

[1-1-1-1- أنواع مدل داده‌ها (Data types) 36](#_Toc107952617)

[2-3-21- چالش‌های مربوط به دسته‌بندی مناسب داده‌های خروجی 42](#_Toc107952618)

[2-3-22- چالش مربوط به اکسل 42](#_Toc107952619)

[2-3-23- چالش‌های مربوط به رابط کاربری و رنگ‌بندی نمودار‌ها 42](#_Toc107952620)

[2-3-24- چالش‌های مربوط به مدل‌سازی و ارتباط بین نمودار‌ها 43](#_Toc107952621)

[2-3-25- چالش‌های مربوط به استفاده از Color Saturation در Tree-map 43](#_Toc107952622)

[2-19- چالش‌های مربوط به تولید کلیدواژه‌ها 43](#_Toc107952623)

[2-20- چالش‌های مربوط به ارزش کلید‌واژه‌ها 43](#_Toc107952624)

[2-21- چالش بدست آوردن فرمول تحلیل‌گر 44](#_Toc107952625)

[2-22- چالش مقایسه نتایج با سایت کدال 44](#_Toc107952626)

[فصل پنجم خروجی نرم‌افزار و تیم تحلیل 46](#_Toc107952627)

[خروجی نهایی و نمودار‌های خروجی 47](#_Toc107952628)

[2-23- نمودار‌های گروهی 47](#_Toc107952629)

[2-3-26- نمودار اول: بررسی مختصر 47](#_Toc107952630)

[2-24- تعداد بازنشر پیام‌های بارگذاری شده 48](#_Toc107952631)

[2-25- تعداد بازدید پیام‌های بارگذاری شده 49](#_Toc107952632)

[2-26- سیگنال‌ دهی انجام شده توسط کانال‌ها 50](#_Toc107952633)

[2-27- تعداد سیگنال‌ها، سیگنال‌های خرید و سیگنال‌های فروش 51](#_Toc107952634)

[2-28- تعداد پاسخ‌های سیگنال‌ دهی شده 51](#_Toc107952635)

[2-29- سیگنال‌ دهی‌هایی که ادیت پیام پنهان‌شده است 51](#_Toc107952636)

[2-3-27- نمودار دوم: بررسی دقیق‌تر 53](#_Toc107952637)

[2-30- گستردگی سیگنال دهی در روز‌های مختلف 53](#_Toc107952638)

[2-31- سیگنال‌دهی انجام‌شده توسط کانال‌ها 53](#_Toc107952639)

[2-32- گستردگی سیگنال خرید و سیگنال فروش 54](#_Toc107952640)

[2-33- تعداد سیگنال خرید/فروش کانال 55](#_Toc107952641)

[2-34- نمودار‌های تکی 57](#_Toc107952642)

[2-3-28- نمودار تکی مربوط به کانال GoldenSignalChannel 57](#_Toc107952643)

[2-3-29- نمودار تکی مربوط به کانال Aryaz\_1 59](#_Toc107952644)

[2-3-30- نمودار تکی مربوط به کانال Bartarinsahm 59](#_Toc107952645)

[2-3-31- نمودار تکی مربوط به کانال BourseSignalTrade 60](#_Toc107952646)

[2-3-32- نمودار تکی مربوط به کانال Bozorgane\_tse 60](#_Toc107952647)

[2-3-33- نمودار تکی مربوط به کانال Hamian\_bourse 61](#_Toc107952648)

[2-3-34- نمودار تکی مربوط به کانال Matinmentalofficial 61](#_Toc107952649)

[2-3-35- نمودار تکی مربوط به کانال Meta\_bourse 62](#_Toc107952650)

[2-3-36- نمودار تکی مربوط به کانال Navasangir100 62](#_Toc107952651)

[2-3-37- نمودار تکی مربوط به کانال TopSignalExchange 63](#_Toc107952652)

[2-3-38- نمودار تکی مربوط به کانال Burs\_signal 63](#_Toc107952653)

[فصل ششم جمع‌بندي و نتيجه‌گيري و پیشنهاداتجمع‌بندي و نتيجه‌گيري 64](#_Toc107952654)

[منابع و مراجع 66](#_Toc107952655)

[Abstract 68](#_Toc107952656)

# فصل اول مقدمهمقدمه

به منظور پایش رویداد‌های فضای مجازی و کشف تقلب در سیگنال دهی می‌بایست نیازمندی‌ها را بررسی کنیم و یک پروژه آرمانی طراحی کنیم. در مراحل بعدی با بررسی‌های بیشتر این پروژه آرمانی به پروژه دست یافتنی نزدیک تر خواهد شد تا مرحله‌ای که پیاده سازی آن امکان پذیر باشد.

## اهداف پروژه

پروژه متشکل از چهار هدف آرمانی، کلی، ویژه و کاربردی می‌باشد.

### اهداف آرمانی

اهداف آرمانی نشان می‌دهد که این پروژه - با بررسی توسعه با زمان هدف اصلی خود – به رسیدن به کدام هدف یا اهداف ایده‌آل کمک می‌کند. در پروژه پایش رویداد‌های فضای مجازی هدف ایده‌آل رسیدن به الگوریتم و روش‌هایی که در کمترین زمان و بهترین کیفیت متقلبین را شناسایی و رفتار آنها را تحلیل کند.

### هدف کلی

هدف کلی پروژه آن چیزی است که با انجام این پروژه می‌توان به آن دست یافت. در پروژه فعلی هدف کلی یافتن راهی به اندازه کافی مناسب به منظور پایش رویداد‌های فضای مجازی می‌باشد.

### اهداف ویژه

اهداف ویژه این پروژه شناخت رویداد‌های فضای مجازی و تعیین روشی کاربردی برای پایش اطلاعات و تحلیل آنها می‌باشد.

### اهداف کاربردی

اهداف کاربردی مشخص می‌کنند که پژوهش در چه زمینه‌هایی و برای چه مصرف کنندگانی کاربرد دارد. پروژه مورد برسی سرمایه‌گزاران را مورد هدف قرار می‌دهد و در زمینه کشف تقلب فعالیت دارد.

## راه‌حل ایده‌آل و نیازمندی‌های مرتبط با پایش رویداد‌های فضای مجازی به منظور نظارت بر بازار سرمایه

اگر تصور کنیم که زمانی نامحدود و منابع نامحدود در دست داشته باشیم چگونه به پیاده‌سازی این پروژه می‌پرداختیم. در مرحله اول می‌بایست فضا‌های مجازی هدف را مورد شناسایی قرار دهیم. در مرحله بعد راه‌حلی ایده آل را پیشنهاد می‌دهیم تا تمامی این فضا‌های مجازی را پوشش دهد.

### فضا‌های مجازی هدف

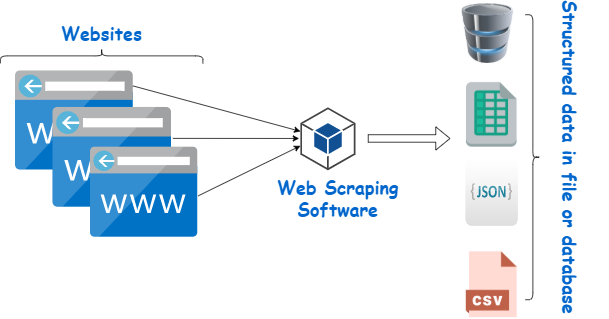
با یک جست‌وجو ساده در گوگل می‌توان سایت‌ها، کانال‌های خبری، گروه‌ها و صفحه‌های زیادی را پیدا کرد که سیگنال دهی انجام می‌دهند. یک پروژه ایده آل تمامی این فضا‌های مجازی را پوشش می‌دهد.

### نیازمندی‌های مرتبط با پایش رویداد‌های فضای مجازی

به‌منظور پایش فضای مجازی می‌توان از برخی مفاهیم برای استخراج و جمع‌آوری داده به نام Web Scraping و WebCrawlingاستفاده کرد.

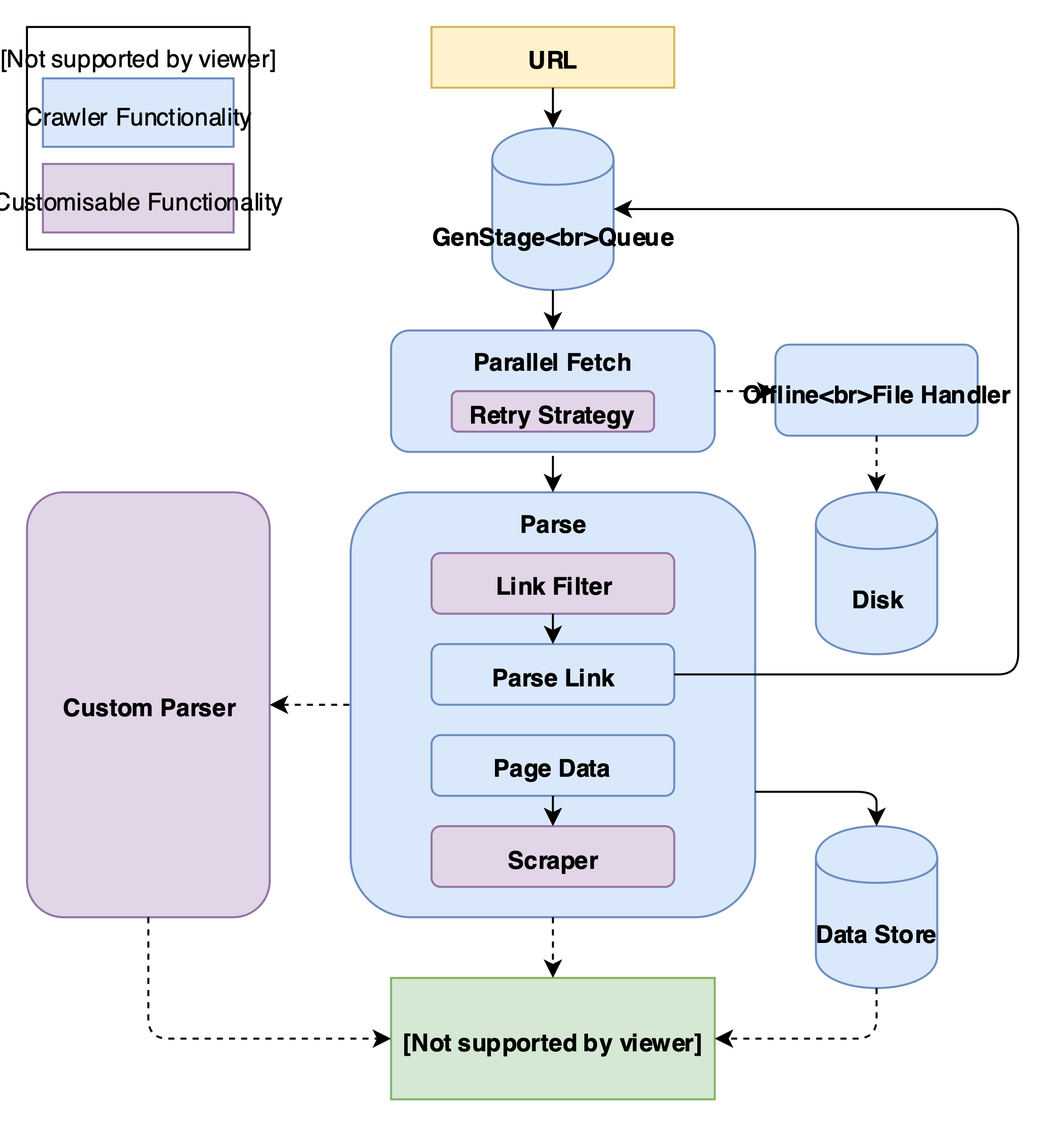
خزنده وب نوعی ربات است که معمولاً توسط موتورهای جستجو مانند گوگل و بینگ اداره می‌شود. هدف آن‌ها ایندکس کردن محتوای وب‌سایت‌ها در سراسر اینترنت است تا آن وب‌سایت‌ها بتوانند در نتایج موتورهای جستجو ظاهر شوند. همچنین خزنده وب به دنبال پیدا کردن دامنه‌ها و لینک‌ها است.

Web Scraping خراش دادن داده‌هایی است که برای استخراج داده‌ها از وب‌سایت‌ها استفاده می‌شود. نرم‌افزار خراش‌دهنده وب ممکن است مستقیماً با استفاده از پروتکل انتقال ابرمتن HTTPیا یک مرورگر وب به وب جهانی دسترسی پیدا کند. درحالی‌که Scraper وب را می‌توان به‌صورت دستی توسط یک کاربر نرم‌افزار انجام داد، این اصطلاح معمولاً به فرآیندهای خودکاری اشاره دارد که بااستفاده از یک ربات یا خزنده وب اجرا می‌شوند. این نوعی کپی است که در آن داده‌های خاصی از وب برای بازیابی یا تجزیه‌وتحلیل آن‌ها در یک پایگاه داده مرکزی جمع‌آوری و کپی می‌شوند.به‌صورت اجمالی می‌توان گفت Web Scraping به معنای استخراج داده‌ها از سایت‌های مختلف است. Web scraping شامل دریافت و استخراج یک صفحه وب می‌شود.



عکس ۲-۱

Web Crawling یکی از مهم‌ترین اجزای تشکیل‌دهنده Web Scraping به‌منظور دریافت کردن یک صفحه برای فرآیند‌های بعدی است. بعد از مرحله دریافت fetchیک صفحه نوبت استخراج مطالب آن می‌رسد. در این مرحله محتوای موجود در یک صفحه وب تجزیه، جستجو و دوباره پیکربندی می‌شود و سپس داده‌های آن در یک پایگاه داده ذخیره می‌شود. Web Scraperها معمولاً داده‌هایی را از صفحات وب استخراج می‌کنند در جای دیگر از آن‌ها را به‌عنوان هدفی دیگر مورداستفاده قرار دهند. برای مثال در اینجا ما می‌خواهیم داده‌هایی را به‌منظور پایش اخبار، اطلاعات و سیگنال‌هایی که در بازار سرمایه داده می‌شود استفاده کنیم.



عکس ۲-۲

با توجه به تعاریف بالا، با استفاده از Web Scraping و Web Crawling می‌توان فضای مجازی را پایش کرد. حال اگر که بخواهیم پایش دقیقی از فضای مجازی داشته باشیم تا بتوانیم بر روی بازار سرمایه نظارت و پیش‌بینی داشته باشیم، باید در دو بخش مشغول به جست‌و‌جو و استخراج داده‌ها شویم.

یکی از داده‌های معتبر به‌منظور استخراج، سایت‌هایی مانند TSETMC است که داده‌های آن را به‌منظور تحلیل می‌توان ذخیره کرد. به‌عنوان Web Scraper می‌تواند در ۳۰ روز متوالی پارامتر‌هایی نظیر خرید، معامله، بازه روز و غیره را پایش کند و اطلاعات آن را در یک پایگاه داده ذخیره کند و سپس در مراحل بعدی از آن ۳۰ روز تحلیلی به دست آید.

بخش دیگر آن مربوط به پایش سیگنال‌ها است. بدین منظور باید Web Scraper داده‌هایی که به نظر می‌رسد تأثیری بر بازار سرمایه دارد را نیز استخراج کند. به‌عنوان‌مثال باید بتواند در برخی از سایت‌های سیگنال دهی جست‌وجو و سیگنال‌های مرتبط را استخراج کند. همچنین اخبار موجود نیز می‌تواند بر بازار سرمایه نقش داشته باشد به‌عنوان‌مثال خبر واردات ۷۰۰۰۰ دستگاه خودروی خارجی یا خبر‌هایی ازاین‌دست، می‌تواند در بازار سرمایه نقش داشته باشد که از این اخبار به‌منظور تحلیل بازار سرمایه استفاده می‌شود و پیش‌بینی‌هایی در این مورد نیز انجام می‌گیرد. بدین منظور نیاز است تا یک لغت‌نامه تهیه شود تا Web Scraper به آن‌ها حساس باشد تا درصورتی‌که مشاهده آن اخبار یا لغت‌ها بتواند آن خبر را ذخیره کند.

پس درنتیجه نیازمندی‌های این بخش به‌منظور پایش فضای مجازی این خواهد بود که در ابتدا یک پایگاه داده باید تهیه شود. سپس با استفاده از زبان‌های برنامه‌نویسی مختلف، یک Web Scraper و یک Web Crawler طراحی ‌شود که به‌صورت ترکیب با یکدیگر کار می‌کنند و با توجه به سایت‌های داده‌شده و همچنین لغت‌نامه داده‌شده، به ترتیب شروع به پایش سایت‌های مرتبط باقیمت گذاری بازار سرمایه و همچنین شروع به پایش برخی از سایت‌های خبری و سیگنال دهی می‌کند، داده‌ها را استخراج می‌کند و سپس موارد استخراج‌شده از این سایت‌ها را در داخل پایگاه داده ذخیره می‌کند تا بخش تحلیل، بتواند از این داده‌ها به‌منظور بررسی و تحلیل بازار سرمایه استفاده کند.

### تحلیل به‌منظور نظارت بر بازار سرمایه

به صورت کلی تمامی اخبار، اطلاعات و پیام‌های پایش شده به دو شکل پیام تصویری و متنی تقسیم می‌شوند.

## تحلیل پیام تصویری

به‌منظور تحلیل پیام تصویری می‌بایست از ابزار‌هایی مانند پردازش تصویر با کمک هوش مصنوعی استفاده کرد. بدین منظور که هر اطلاعات پایش شده تصویری که توسط بخش پایش اطلاعات، در پایگاه داده موجود است را پردازش و تحلیل کنیم. به‌عنوان‌مثال پستی در شبکه مجازی اینستاگرام پایش شده است که در آن سیگنال‌ فروش سهم خاص داده‌شده است. این اطلاعات می‌تواند مفید باشد و به کمک آن‌ها کشف تقلب صورت بگیرد.

## تحلیل پیام متنی

به‌منظور تحلیل پیام متنی می‌بایست از ابزار‌هایی مانند پردازش زبان طبیعی یا NLP استفاده کنیم. به‌طور خاص‌تر از تکنیک NLP Labeling استفاده می‌کنیم. با استفاده از NLP، معامله‌گران و سرمایه‌گذاران می‌توانند تشخیص دهند که آیا مفهوم کلی بیان‌شده در گزارش سالانه مثبت است، منفی است، یا هیچ‌کدام و غیره. نتیجه این تحلیل مفهوم کلی می‌تواند برای تصمیم‌گیری معاملات مورداستفاده قرار گیرد. NLP همچنین برای تجزیه‌وتحلیل رونوشت‌های تماس‌های کسب درآمد برای شناسایی سیگنال‌ها استفاده می‌شود. در ادامه با برسی سیگنال دهی انجام شده و مقایسه نتایج با وضعیت بازار می‌توانیم تحلیل خوبی را انجام دهیم و افرادی که مستعد به تقلب هستند را شناسایی کنیم.

با استفاده از این تکنیک سایت‌های موجود در ایندکس دیتابیس گردآوری‌شده را می‌توانیم تحلیل و نتیجه‌گیری ‌کنیم. مرحله بعد کشف معنا و مفهوم با decode کردن syntax است که برای این کار به NLP نیاز داریم. به‌عنوان یک تعمیم گسترده، می‌توان گفت که NLP اساساً یعنی گرفتن یک سند غیر شفاف، متشکل از مجموعه منظمی از نمادها که به یک دستور زبان کاملاً تعریف‌شده به نحو مناسب پایبند هستند، و درنهایت استنتاج کردن معنا و مفهوم مرتبط با آن نمادها. اولین گام در اکثر خطوط لوله‌های (Pipeline) پردازش زبان طبیعی در تشخیص جملات، ممکن است فقط شمردن نقطه‌ها، علامت‌های سؤال و علامت تعجب در جمله باشد. این واضح‌ترین ابتکار (heuristic) برای شروع است، اما کاملاً خام است و پتانسیل ایجاد خطای بسیار بالایی را دارد.

توانایی انجام و اجرای یک تشخیص نسبتاً خوب از جملات به‌عنوان بخشی از رویکرد NLP برای استخراج داده‌های بدون ساختار، می‌تواند برخی از قابلیت‌های متن‌کاوی بسیار قدرتمند، مانند تلاش‌های خام اما بسیار منطقی برای خلاصه‌سازی اسناد را برای ما امکان‌پذیر سازد؛ و احتمالات و رویکردهای متعددی وجود دارد. ما می‌بایست شباهت در متن را بیابیم، اطلاعات را استخراج‌کنیم، رابطه میان معنا و مفهوم متن را بیابیم تا درنهایت خلاصه‌ای منسجم و روان شامل تمام نکات اصلی متن را به دست آوریم.

# فصل دوم مشخص کردن محدوده و ساده‌سازی پروژه

# تفاوت پروژه ایده‌آل و پروژه هدف

نیازمندی‌های مشخص‌شده و اطلاعات جمع‌آوری‌شده مربوط به یک پروژه ایده‌آل می‌باشد و باوجود زمان محدود و منابع محدود فعلی امکان‌پذیر نمی‌باشند. درنتیجه در جلسات مطرح‌شده بین تیم توسعه محصول و سهام‌داران پروژه به ساده‌سازی و مشخص کردن محدوده پروژه پرداختیم.

## مشخص کردن Scope و Not Scope

یکی از مهم‌ترین قسمت‌های هر پروژه مشخص کردن محدوده پیاده‌سازی و موارد خارج محدوده هست. درصورتی‌که این مرحله به‌خوبی صورت نگیرد پروژه در انتها با شکست روبه‌رو خواهد شد.

طی جلسات پی‌درپی تیم توسعه محصول و سهام‌داران پروژه (استاد درس) محدوده پروژه دچار تغییراتی شد. همچنین برخی از این تغییرات در حین پیاده‌سازی رخ‌داده. به ترتیب به معرفی محدوده و خارج محدوده می‌پردازیم.

### محدوده پروژه

محدوده پروژه شامل موارد ذیل می‌باشد:

* جمع‌آوری اطلاعات و پایش رویداد‌ها مربوط به شبکه اجتماعی تلگرام می‌باشد.
* تنها محتوای متنی مورد بررسی قرار می‌گیرد.
* پیام‌های یک الی هفت روز گذشته نیاز به تحلیل دارند.
* پیام‌های کانال‌های تلگرامی مورد بررسی قرار گیرند.
* پیام‌های کانال‌های پربازدید مورد بررسی قرار گیرند.
* تحلیل پیام‌ها با استفاده از نماد‌ها صورت گیرد.
* تحلیل پیام‌ها با استفاده از تعداد بازدید و ارسال صورت گیرد.
* خروجی تحلیل‌گر مشخص‌کننده سیگنال دهی و یا عدم سیگنال دهی باشد.
* خروجی تحلیل‌گر مشخص‌کننده سیگنال خرید یا فروش باشد.
* از زبان برنامه‌نویسی پایتون در پیاده‌سازی استفاده شود.
* از فریمورک جانگو در پیاده‌سازی استفاده شود.
* نتایج تحلیل‌گر در قالب نمودار نمایش داده شود.

### خارج محدوده پروژه

محدوده پروژه شامل موارد ذیل نمی‌باشد.

* جمع‌آوری اطلاعات از شبکه‌های اجتماعی اینستاگرام، توییتر، واتس‌آپ، رددیت و غیره نمی‌باشد.
* محتوای تصویری مورد بررسی قرار نمی‌گیرد.
* پیام‌های قدیمی‌تر مورد بررسی قرار نمی‌گیرد.
* برای تحلیل از هوش مصنوعی کمک گرفته نمی‌شود.
* در تحلیل اطلاعات فرستنده پیام استخراج نمی‌شود.
* از زبان‌های برنامه‌نویسی جاوا، سی، سی پلاس پلاس، سی شارپ، گولنگ و غیره استفاده نمی‌شود.
* خروجی تحلیل‌گر کشف تقلب انجام نمی‌دهد.

## معرفی نیازمندی‌های شناسایی‌شده در قالب Product Backlog

کاربرد متدولوژی‌های چابک در پروژه‌هایی که ایده اصلی مشخص ولی جزئیات نامشخص هستند بسیار پرکاربرد است. ما نیز به همین دلیل از متدولوژي‌های چابک استفاده می‌کنیم.

برای این منظور ابتدا خواسته‌ها را مشخص می‌کنیم. خواسته‌های واقعی‌تر و قابل پیاده‌سازی در شرایط فعلی پروژه را انتخاب می‌کنیم و آن‌ها را خواسته می‌نامیم. درنهایت خواسته‌هایی که در قالب بازه زمانی دو الی چهار هفته قابل پیاده‌سازی هستند را انتخاب می‌کنیم و در قالب نیاز‌ها مطرح می‌کنیم. درنهایت نیاز‌ها را به Product Backlog اضافه می‌کنیم و پیاده‌سازی می‌کنیم.

بعد از جلسات مکرر با اعضا تیم توسعه و سهام‌داران Product Backlog پیاده‌سازی شد که به توضیح آن می‌پردازیم. ابزار مورداستفاده ما نرم‌افزار تحت وب ماکروسافت Azure هست. در مرحله اول Epic‌ها را مشخص می‌کنیم. هر Epic به تعداد زیادی Feature شکسته می‌شوند. Featureها نیز به Product Backlog شکسته می‌شوند که افزونه‌های اصلی پروژه می‌باشند. باید در تعریف Product Backlog حواسمان باشد که تمامی موارد در دوره دو الی چهار هفته‌ای قابل پیاده‌سازی باشند. در غیر این صورت به موارد کوچک‌تر تقسیم شوند.

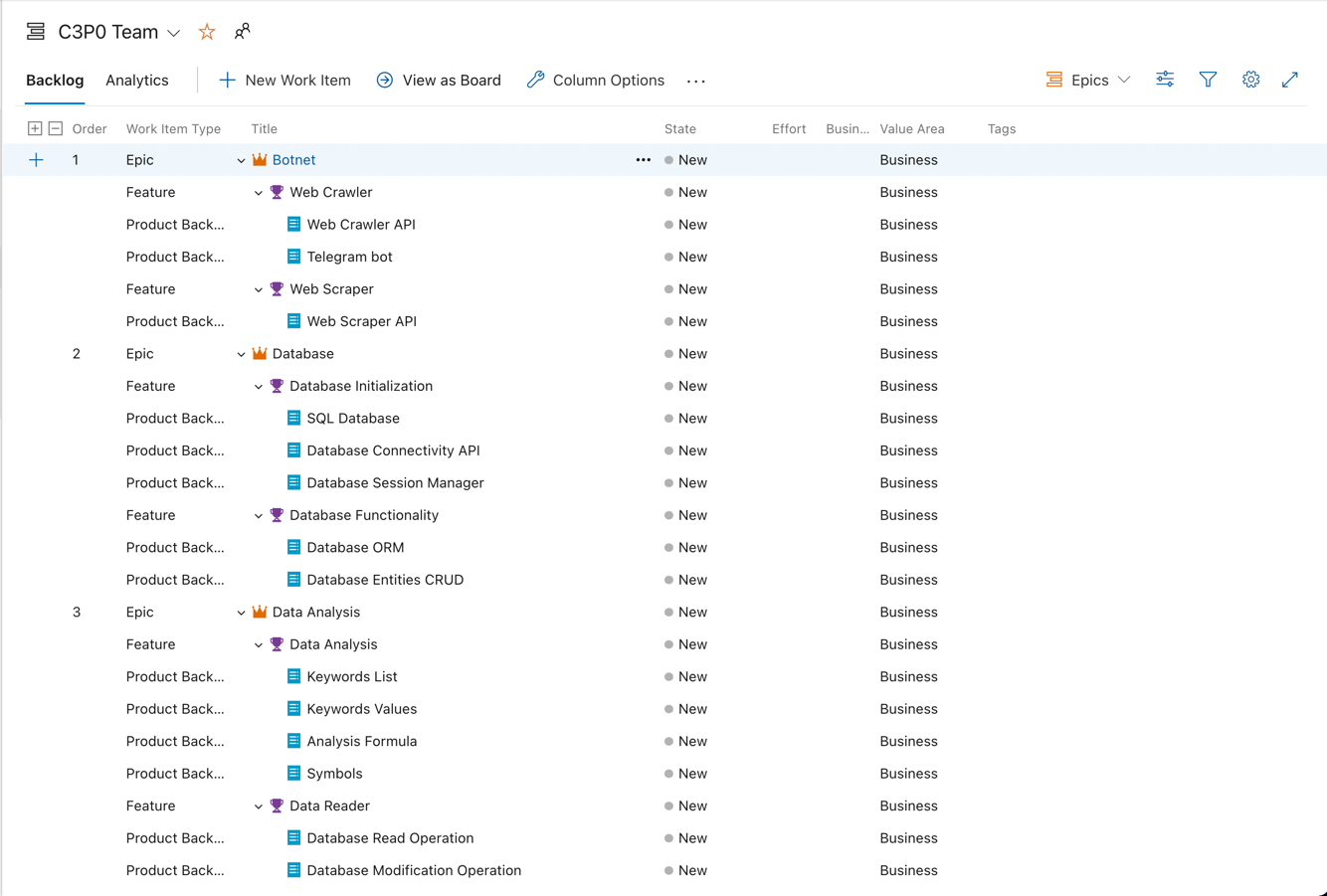
### Product Backlog مربوط به تیم توسعه محصول

تیم توسعه محصول در پیاده‌سازی نرم‌افزار پروژه نقش دارد. این پروژه به صورت یک نرم‌افزار تحت وب پیاده‌سازی شده است و از فریمورک جنگو استفاده می‌کند. استفاده از این فریمورک بسیار مفید است چراکه از کتابخانه‌های بسیار گسترده و مفید زبان پایتون استفاده می‌کند.

PBIهای مربوط به تیم توسعه محصول (عکس ۲-۳) در سه دسته بات‌نت، پایگاه‌داده و تحلیل‌گر تقسیم می‌شود. ماژول بات‌نت با کمک یک کلید و اتصال به تلگرام پیام‌های کانال تلگرامی از پیش مشخص شده را که در ۲۴ ساعت گذشته پیامی را به اشتراک گذاشته باشند می‌خواند و در پایگاه داده قرار می‌دهد.

ماژول پایگاه داده مربوط به‌تمامی کاربرد‌های پایگاه داده و پیاده‌سازی توابع پایگاه داده می‌باشد. در این پروژه از پایگاه داده Sqlite3 استفاده‌شده است.

ماژول تحلیل‌گر با کمک نماد‌هایی که در اختیار دارد اطلاعات را از پایگاه داده می‌خواند و پیام‌ها را تحلیل می‌کند. دوباره تحلیل‌های صورت گرفته را به پایگاه داده برمی‌گرداند تا تیم تحلیل‌گر از آن‌ها در ساختن نمودار‌ها استفاده کنند.



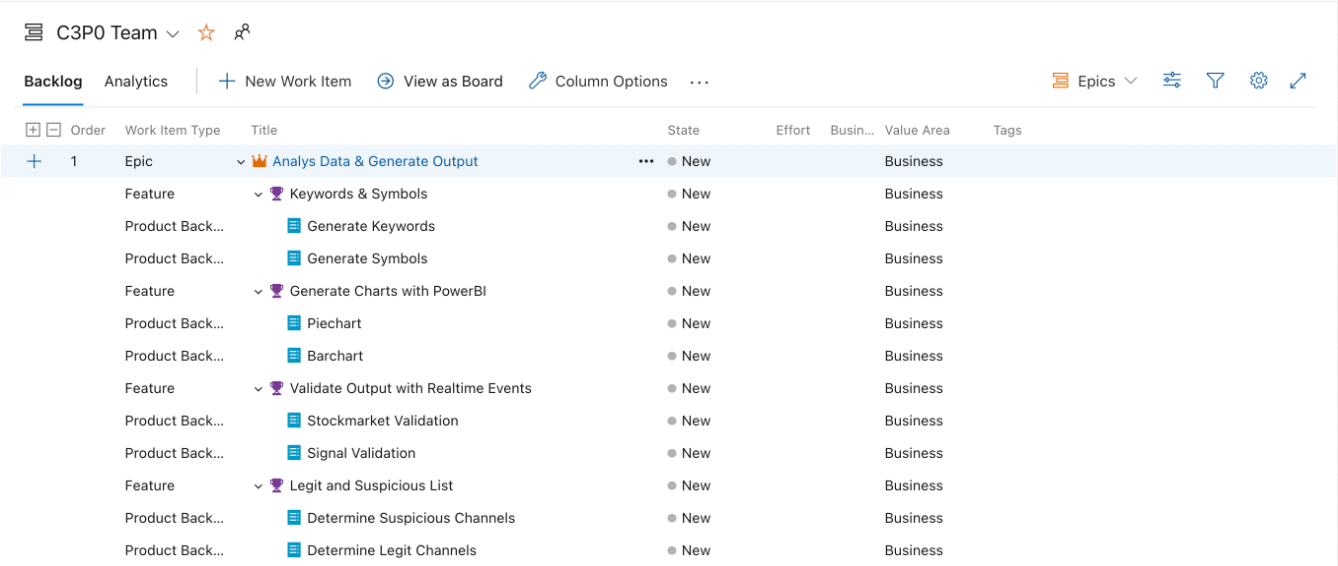
عکس ۲-۳

### تیم تحلیل‌گر

تیم تحلیل‌گر با استفاده از خروجی تحلیل‌شده تیم توسعه محصول نمودار‌هایی را تشکلی داده و آن‌ها را تحلیل می‌کنند. درصورتی‌که در تحلیل صورت گرفته رفتار مشکوکی مشاهده شود و یا تقلبی کشف شود این تقلب گزارش داده خواهد شد.

تیم تحلیل‌گر با دریافت اطلاعات تحلیل‌شده توسط تیم توسعه و تطابق آن با رویداد‌های بورس اوراق بهادار کشف تقلب را انجام می‌دهند. این عملیات به این صورت است اگر کانال تلگرامی وجود داشته باشد که بازدید بالایی دارد و در اکثر موارد سیگنال دهی درستی انجام می‌دهد ممکن است اطلاعات نهانی - هرگونه اطلاعات افشاء نشده برای عموم که به‌طور مستقیم و یا غیرمستقیم به اوراق بهادار، معاملات یا ناشر آن مربوط می‌شود و در صورت انتشار بر قیمت و یا تصمیم سرمایه‌گذاران برای معامله اوراق بهادار مربوط تأثیر می‌گذارد – داشته باشد. این صرفاً حدس است و نیاز به بررسی‌های عمیق‌تر دارد.

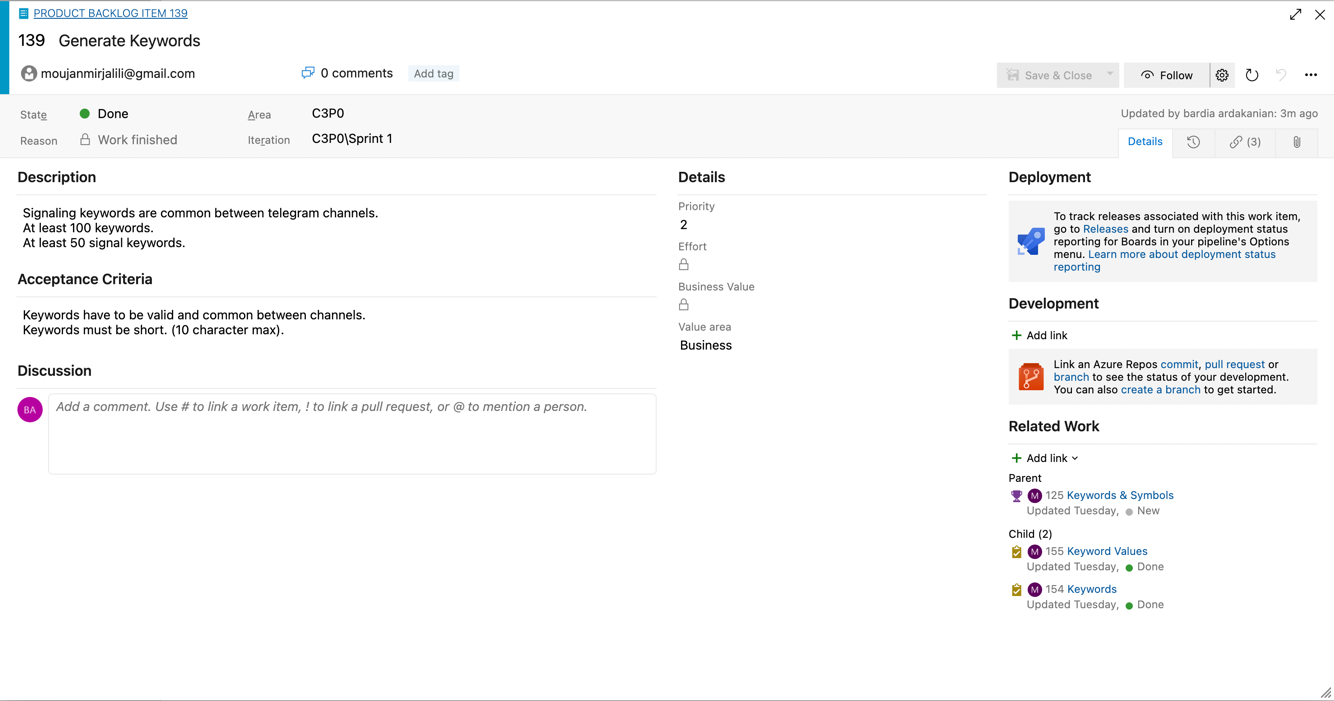
PBIهای تیم تحلیل (عکس ۲-۴) تنها در آنالیز کردن و تولید خروجی خلاصه می‌شود. این تحلیل با کمک متخصصین صورت می‌گیرد و ربات‌ها.



عکس ۲-۴

### تعریف مفاهیم Acceptance Criteria و Definition of Done

برای هر Product Backlog Item مفاهیم مربوطه تعریف شده‌اند. این مفاهیم با کلیک کردن بر روی هر مورد قابل مشاهده هستند (عکس ۲-۴). تمامی این مفاهیم لیست شده و قابل مشاهده هستند.



عکس ۲-۵

## تعریف مفاهیم برای هر Product Backlog Item

* Telegram bot
  + DoD: Bot can connect and read data from telegram.
  + AC: Data must have columns such as published date, views, forwards count, message id and channel name.
* Web crawler
  + DoD: Crawl data from telegram channels and store them in database.
  + AC: Must be text-only content. A message cannot be empty. Old messages are invalid (30 day old at max).
* SQL Database
  + DoD: Initialize database and make connectivity module.
  + AC: Database must be accessible through its module.
* Database connectivity API
  + DoD: API to access database remotely.
  + AC: APIs must have security filters to be considered safe.
* Keywords list
  + DoD: Read stores keywords from database (CRUD operation).
  + AC: All keywords must be cached and used in analysis.
* Keywords value
  + DoD: Assign values to keywords and store them in database.
  + AC: Values must be group-specific unique in order to distinguish signals from one another
* Analysis Formula
  + DoD: Formula for determining signaling, buy signaling, sell signaling from one another.
  + AC: Formula must be normalized and has a precision of 95%.
* Database read operation
  + DoD: Read from Message, Symbol, Keywords, Keyword-values tables.
  + AC: Read operation must be atomic. Redis must be used as session manager.
* Database modification operation
  + DoD: Update rows in Message, Symbol, Keywords, Keyword-values tables.
  + AC: Update operation must be atomic. Redis must be used as session manager.
* Generate Symbols
  + DoD: Extract all Symbols from TSETMC.
  + AC: Symbols need code, group, industry, table, English symbol, English name, symbol, name columns.
* Pie Chart
  + DoD: Pie chart of all distribution entities.
  + AC: Pie charts has to be color coded and easy to understand.
* Bar Chart
  + DoD: Bar chart of all attendance entities.
  + AC: Pie charts has to be color coded and easy to understand.
* Signal Validation
  + DoD: Validate signal output from database.
  + AC: Most (80%) generated signals must be correct. Otherwise need better formula
* Stock-market Validation
  + DoD: Validate signal with real-time stock-market events.
  + AC: Extracted signals should be comparable with real-time stock-market.
* Determine suspicious channels
  + DoD: Compare extracted output with real-time events and determine suspicious channels.
  + AC: Channels must violate a set of rules in order to be considered suspicious.
* Determine Legit channels
  + DoD: Compare extracted output with real-time events and determine Legit channels.
  + AC: Legit channels cannot violate any rules.

## برنامه‌ریزی اسپرینت‌ و جلسات اسکرام

مهم‌ترین رکن یک تیم موفق توانای برنامه‌ریزی اهدافش در مدت زمانی مشخص است. برای پیاده‌سازی پروژه سه اسپرینت خواهیم داشت که هر اسپرینت متشکل از ۱۲ روز کاری می‌باشد. با احتساب روز‌های تعطیل بین اسپرینت‌ها هر اسپرینت ۲ هفته طول می‌کشد.

### مکان جلسات

طبق توافق اعضا تیم تمامی جلسات در سایت دانشکده مهندسی کامپیوتر دانشگاه امیرکبیر صورت گرفته‌اند. مکان جلسات همیشه ثابت بوده و گوشه‌ای از سایت می‌باشد. زمان اجرای جلسات نیز در تمامی مراحل یکسان بوده و به‌هیچ‌وجه تغییر نکرده‌اند.

جلسات Daily به مدت ۱۵ دقیقه روز‌های زوج رأس ساعت ۹ صبح اجرا شدند. جلسات Review به مدت دو ساعت در انتهای هر اسپرینت ساعت ۹ الی ۱۱ برگزار شدند. جلسات Planning بلافاصله بعد از جلسه Review به مدت یک ساعت و ۲۰ دقیقه برگزار شدند. با اتمام جلسات فوق جلسه Retrospective را خواهیم داشت که مدت زمان ۳۰ دقیقه طول می‌کشند. مکان تمامی جلسات سایت دانشکده می‌باشد.

### مراحل اتمام هر Product Backlog

هر PBI که در اسپرینت آورده می‌شود توسط تیم توسعه محصول به تعدادی Task شکسته می‌شود. هر Task مراحل ذیل را طی می‌کند:

* To Do
* In Progress
* Technical Review
* Test
* Done

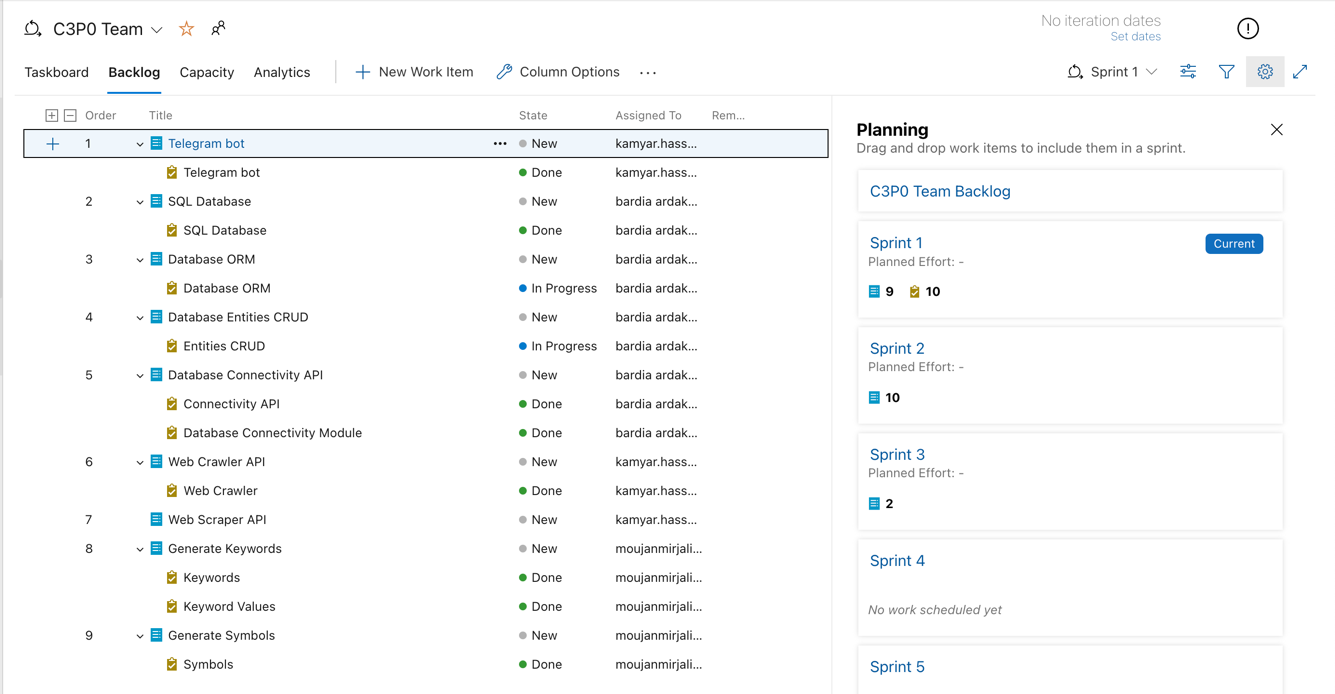
هر وظیفه در ابتدا در وضعیت To Do قرار دارد و بعدازاینکه توسعه‌دهنده شروع به کار کردن می‌کند به وضعیت In Progress وارد می‌شود. بعدازاینکه توسعه‌دهنده وظیفه را به اتمام برساند وضعیت آن را به Technical Review تغییر می‌دهد تا مدیر پروژه وظیفه را بررسی کند و درصورتی‌که با Definition Of Done تطابق نداشته باشد به توسعه‌دهنده اطلاع دهد و وضعیتش را به To Do تغییر دهد.

درصورتی‌که تائید مدیر پروژه را داشته باشد به وضعیت Test وارد می‌شود تا توسط تیم تحلیل آزموده شود. با اتمام آزمایش ها تیم تحلیل به وضعیت اتمام یا Done وارد می‌شود و بر افزونه‌ها پروژه اضافه می‌شود. درنهایت توسعه‌دهنده سراغ وظیفه دیگری می‌رود ولی آخر.

### اسپرینت اول پیاده‌سازی پایگاه‌داده، بات‌نت و استخراج نماد‌ها.

در اسپرینت اول بیشتر کار‌ها مربوط به تیم توسعه محصول می‌باشد. در این اسپرینت PBI‌های انتخاب و پیاده‌سازی شدند. همچنین تیم تحلیل به دلیل اینکه تیم توسعه برای استخراج و تحلیل اولیه پیام‌ها نیازمند نماد‌ها، کلیدواژه‌ها و ارزش کلیدواژه‌ها بودند به جمع‌آوری این موارد پرداختند.

مواردی که در این اسپرینت به آن‌ها پرداخته شد و وظایفی که به آن‌ها شکسته شدند در عکس ۳-۱ قابل‌مشاهده است.

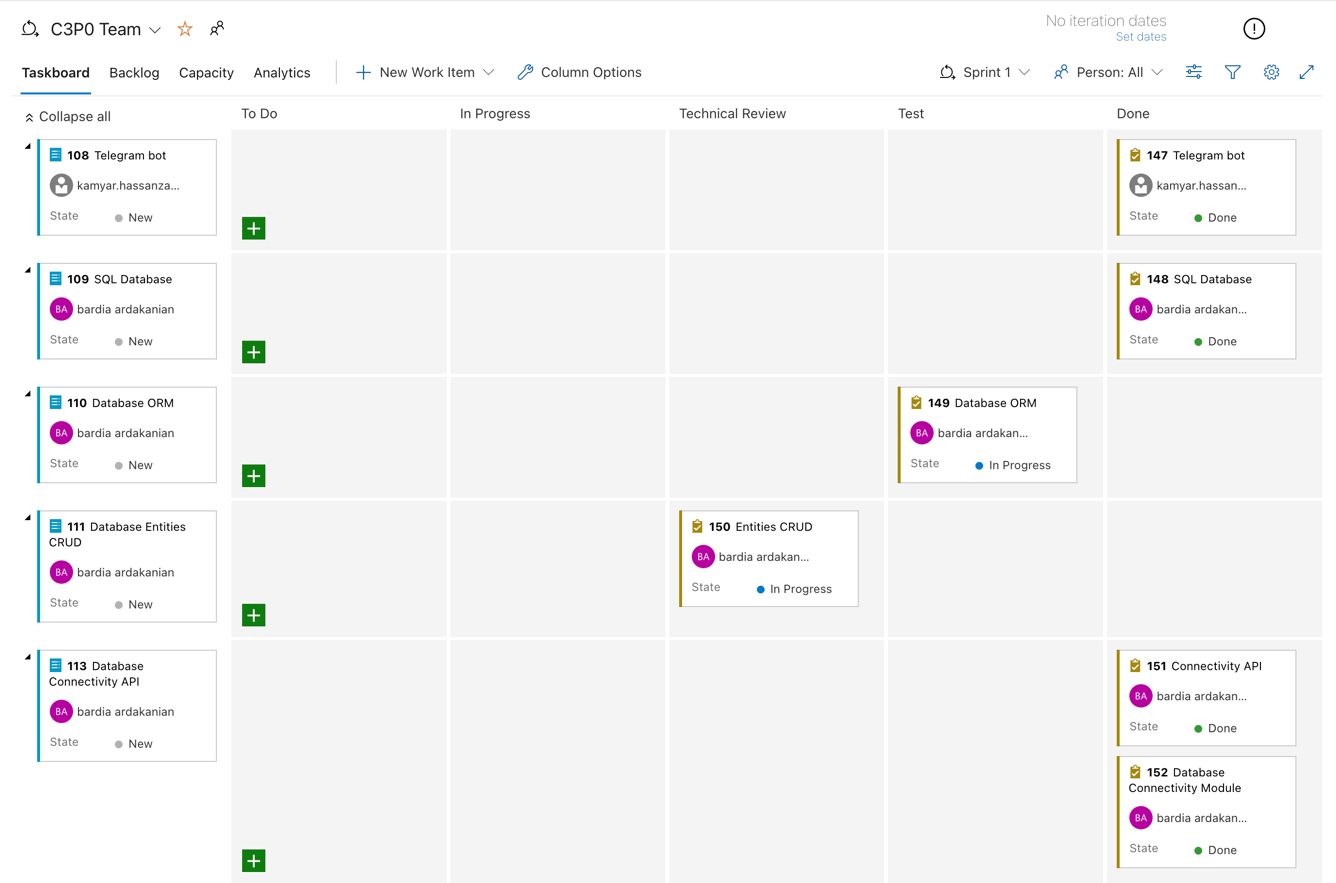


عکس ۲-۶

در بین وظایف، طبق بررسی‌های صورت گرفته وظیفه Web Scraper API خارج از محدوده پروژه بود و حذف گردید. (عکس ۳-۲) همچنین دو وظیفه در وضعیت تست و بررسی تکنیکال مانده بودند که به اسپرینت بعدی فرستاده شدند. بقیه وظایف در این اسپرینت به پایان رسیدند.

در جلسه Review به بررسی عملکرد تیم توسعه پرداختیم. در این بررسی به نتیجه رسیدیم که بیشتر از ظرفیت تیم وظیفه اختصاص داده‌ایم و این باعث کاهش بازدهی افراد شده است. همچنین تخمین زمان اختصاص‌یافته به وظایف غیرواقعی و از مقدار واقعی کمتر بوده‌اند. در اسپرینت بعدی این موارد اصلاح خواهند شد.

نتیجه جلسه Planning اسپرینت دوم در بخش‌ بعدی اعلام شده است. در جلسه Retrospective عوامل موفقت شناخته شدند و در اسپرینت بعدی به کار گرفته شدند. همچنین راه‌حل‌های جدیدی پیشنهاد داده شدند.



عکس ۲-۷

## عوامل پیروزی در اسپرینت اول: جلسات روزانه

جلسات روزانه بیشتر از آنچه انتظار داشتیم نیاز به جلسات دیگر را کاهش داد. درنتیجه این بررسی تصمیم گرفته شد تا جلسات روزانه بجای روز‌های زوج به هرروز تغییر پیدا کند.

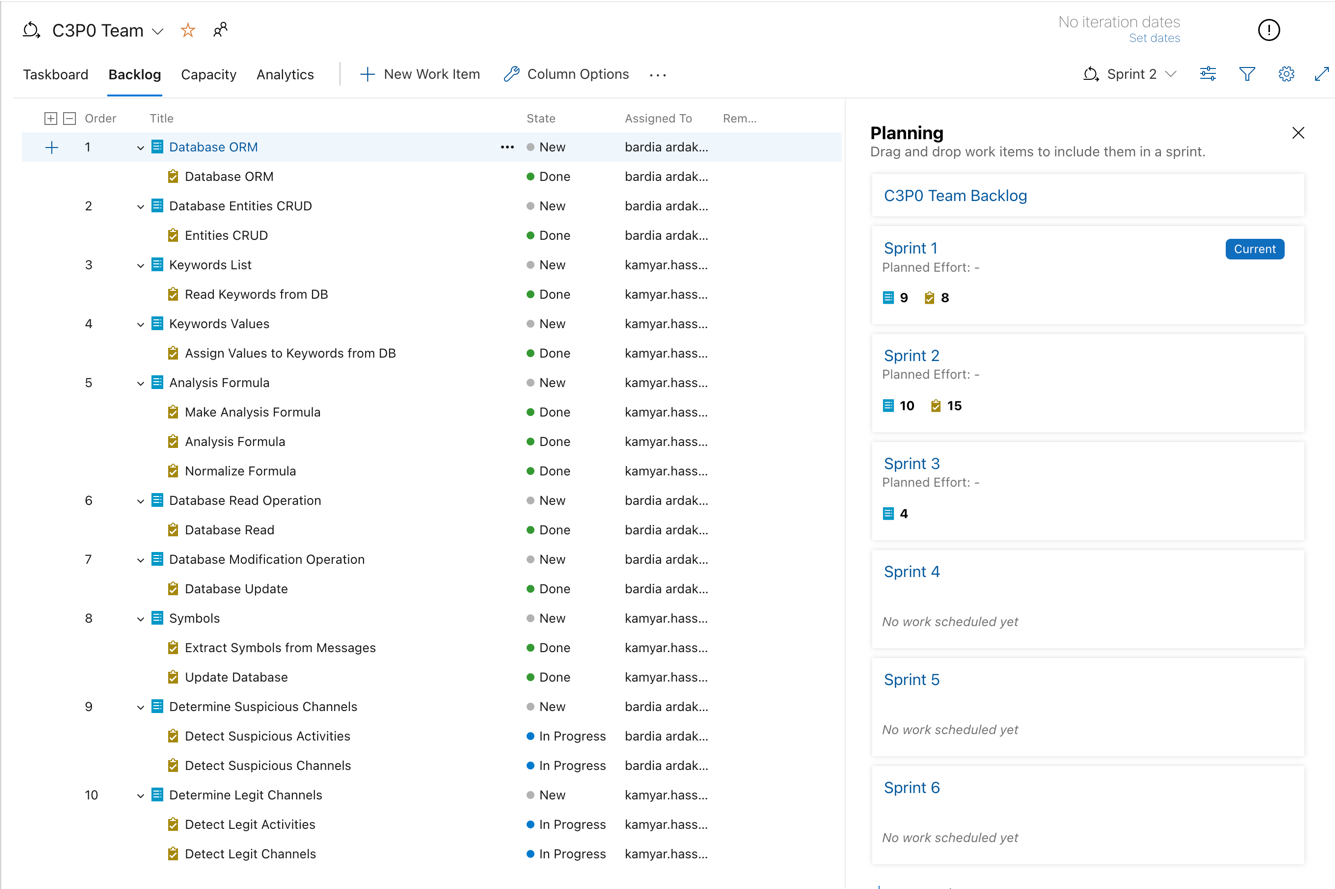
## عوامل پیروزی در اسپرینت اول: کار کردن در محیط اشتراکی

در اسپرینت اول یکی از عوامل موفقیت کار کردن در کنار یکدیگر بود که باعث بالا رفتن بازدهی افراد تیم شد. تلاش کردیم تا در اسپرینت‌های بعدی نیز از این کیفیت استفاده کنیم. به همین جهت ساعاتی در روز (معمولا ۱۵ الی ۱۸) در سایت دانشکده و یا سالن مطالعه دانشکده دورهم جمع شدیم و به پیاده‌سازی پرداختیم.

### اسپرینت دوم پیاده‌سازی تحلیل‌گر و تحلیل متخصصین

در اسپرینت دوم با بهره‌برداری از تجربیات اسپرینت اول، برنامه‌ای منطقی‌تر برای اسپرینت دوم در نظر گرفتیم. همچنین وارد فاز دوم پیاده‌سازی یعنی تحلیل شدیم. در این اسپرینت ابتدا تابعی پیاده‌سازی کردیم که در ادامه با کارکرد آن به‌طور دقیق آشنا خواهیم شد. این تابع با کمک ارزش نماد‌ها امتیازی به هر پیام داده و سیگنال بودن/نبودن و سیگنال خرید/فروش بودن را تشخیص می‌دهد.

برای بدست آوردن فرمول ذکرشده نرمال‌سازی‌هایی صورت گرفت که به بهترین فرمول ممکن دست‌یابیم. مواردی که در این اسپرینت به آن‌ها پرداخته شد و وظایفی که به آن‌ها شکسته شدند در عکس ۳-۳ قابل‌مشاهده است.

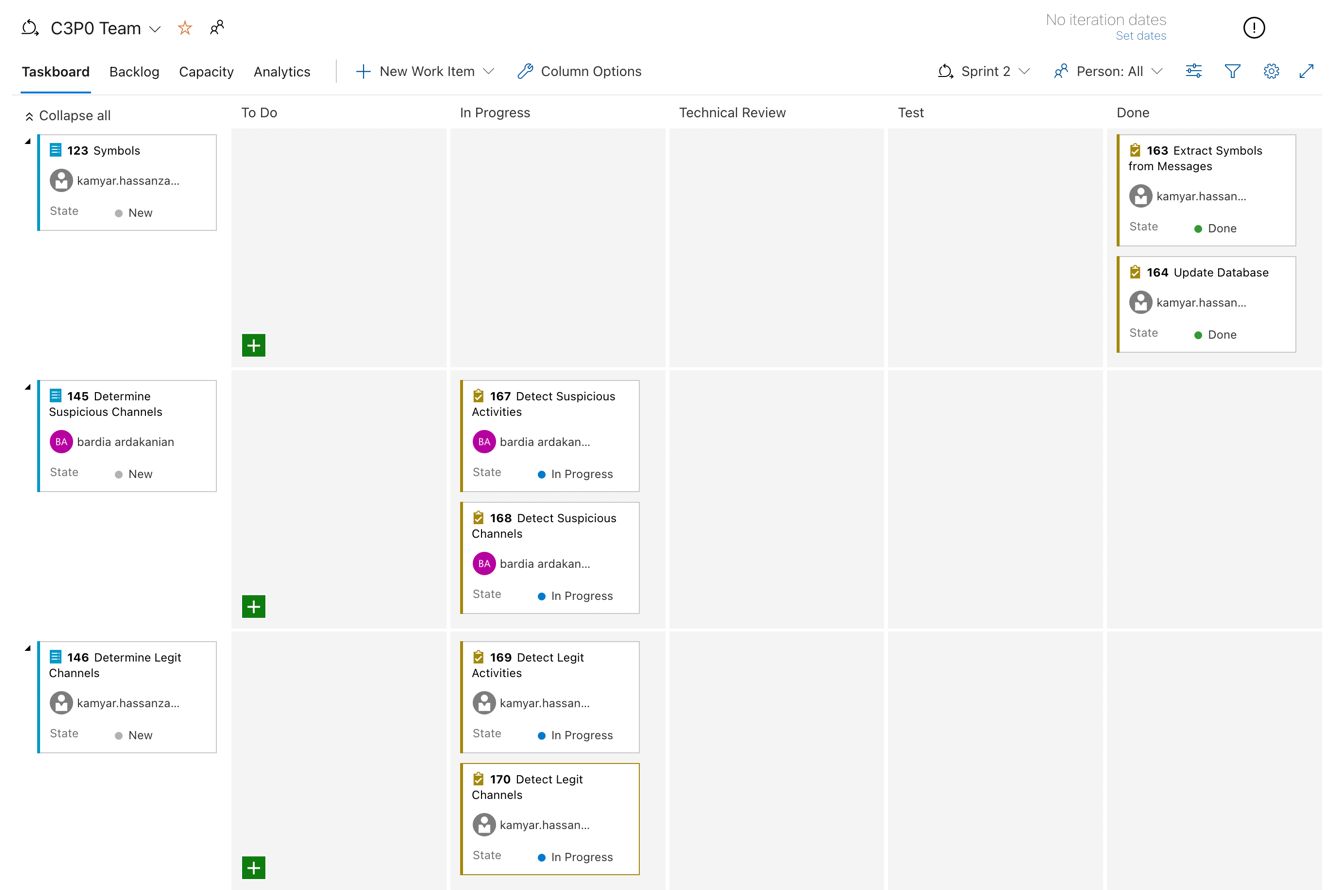


عکس ۲-۸

با توجه به پیچیدگی تشخیص مشکوک و غیر مشکوک بودن کانال‌ها این وظیفه تکمیل نشد (عکس ۳-۴) و به انتها نرسید. بنابر‌این به اسپرینت بعد انتقال داده شد تا در اسپرینت بعدی موردبررسی قرار گیرد.

در جلسه Review به بررسی عملکرد تیم توسعه و تحلیل پرداختیم. در این بررسی به نتیجه رسیدیم که اختصاص وظایف باکیفیت بهتری انجام‌شده و همین روند باید ادامه یابد.

نتیجه جلسه Planning اسپرینت سوم در بخش‌ بعدی اعلام‌شده است. در جلسه Retrospective به بررسی عوامل موفقیت پرداختیم و به نتیجه رسیدیم که عملکرد خوبی داشتیم و عوامل موفقیت شناخته‌شده در اسپرینت اول را باید ادامه دهیم. همچنین به پیشنهاد یکی از افراد تیم توسعه برای توسعه نرم‌افزار روز‌هایی را به باغ کتاب رفتیم تا در کنار هم و سکوت بیشتر به فعالیت بپردازیم که نتیجه بسیار مطلوبی داشت.

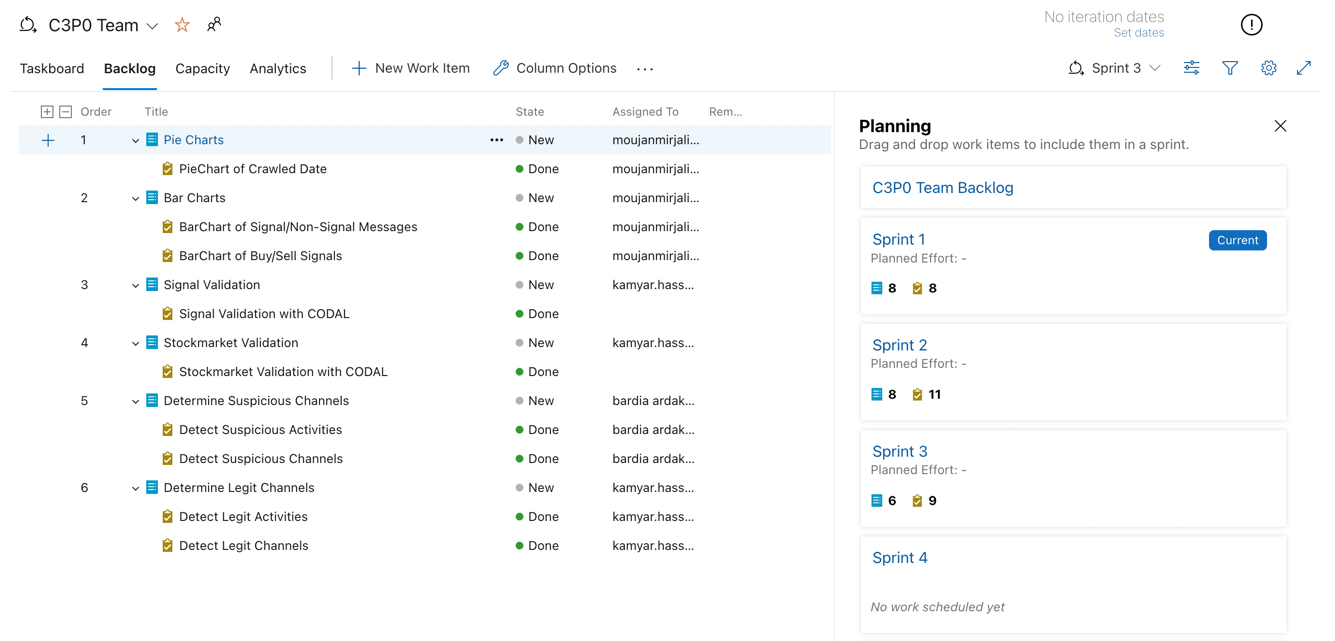


عکس ۲-۹

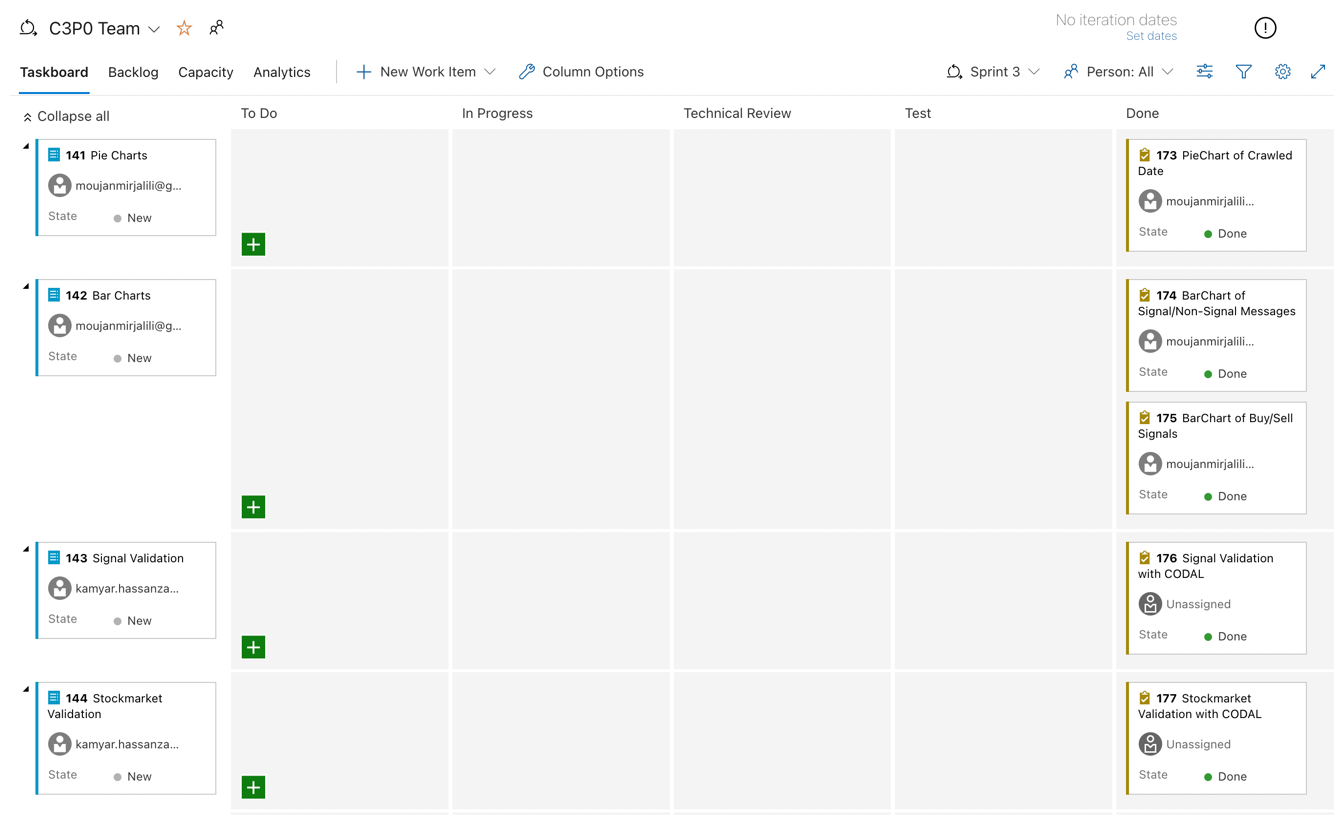
### اسپرینت سوم تحلیل و خروجی نهایی

طبق جدول زمان‌بندی اسپرینت سوم اسپرینت آخر پروژه می‌باشد که در آن می‌بایست نرم‌افزاری کاربردی داشته باشیم و بتوانیم داده‌هایی مفید را از آن استخراج‌کنیم. همچنین در این اسپرینت می‌بایست داده‌های استخراج‌شده و نمودار‌های ایجادشده را تحلیل کنیم و با اخبار بورس اوراق بهادار مقایسه کنیم. درصورتی‌که رفتاری مشکوک مشاهده کنیم گزارش می‌دهیم. به‌عنوان‌مثال اگر کانالی به‌صورت عمومی سیگنال دهی درست انجام دهد و بعضی از این سیگنال دهی‌ها برخلاف شواهد فعلی بازار بورس اوراق بهادار باشد به‌احتمال‌زیاد اطلاعات نهانی دارد و از این اطلاعات سوءاستفاده می‌کند. این تخلفات باعث از بین رفتن اعتماد میان سهام‌داران می‌شود و باید با افرادی ازاین‌دست برخورد شود.

برای محقق شدن این عمر چندین وظیفه تعریف و میان اعضا تیم تقسیم شد. این وظایف در عکس ۳-۵ قابل‌مشاهده هستند.



عکس ۲-۱۰



عکس ۲-۱۱

با نتایج حاصل‌شده توسط نمودار‌ها و طبق این اطلاعات با سایت CODAL توانستیم کانالی را به‌عنوان کانال مشکوک نشانه‌گذاری کنیم. البته این بررسی باید عمیق‌تر باشد و نمی‌توان نتیجه قطعی بدست آورد.

در جلسه Review به بررسی عملکرد تیم توسعه و تحلیل پرداختیم. در این بررسی به نتیجه رسیدیم که عملکرد ضعیف‌تری داشتیم و بازدهی تیم توسعه و تحلیل نزدیک به بیست درصد کاهش داشته است. البته مواردی همچون بازه امتحانات در این موضوع تأثیر بسیاری داشته است.

در جلسه Retrospective به بررسی عوامل موفقیت پرداختیم و به نتیجه رسیدیم که برای بهبود بازدهی می‌بایست ساعاتی مشخص و روتین را در روز به پیاده‌سازی پروژه اختصاص دهیم. ساعات شناور بسیار تأثیر نامطلوبی داشته‌اند.

# فصل سوم چالش‌های تیم توسعه محصول

# ایده اولیه

در ابتدا پیاده‌سازی تیم توسعه قصد داشت تا با کمک خزنده وب اطلاعات را از سایت‌های مختلف من‌جمله کانال‌های تلگرامی جمع‌آوری کند و در پایگاه داده MongoDB ذخیره کند. در محله‌های بعدی این اطلاعات را از پایگاه داده برداشته، گزینش کند و منتخبین را در پایگاه داده Sqlite3 ذخیره کند. این کاوش اطلاعات و استخراج از پایگاه داده اول به‌صورت موازی و با کمک یک صف کاری (Job Queue) پیاده‌سازی شود.

همچنین در ابتدا خبری از فریمورک جنگو نبود و به پیاده‌سازی به‌صورت خالص با پایتون بسنده کرده بودیم. در ادامه کار با چالش‌هایی روبه‌رو شدیم که نیازمند تغییر در ساختار پیاده‌سازی پروژه بود.

## چالش‌های مربوط به پایگاه داده

به علت دوباره‌کاری شدن و استفاده نکردن بهینه از فضا و منابع بجای استفاده از دو پایگاه داده تنها از یک پایگاه داده استفاده کنیم. در این حالت مشکلاتی من‌جمله موازی‌سازی عملیات استخراج پیام و تحلیل نخواهیم داشت چراکه کماکان می‌توانند به‌صورت موازی از هم فعالیت کنند.

## چالش‌های مربوط به‌صف کار

پیاده‌سازی این بخش خارج از محدوده پروژه بود و وقت و انرژی زیادی از تیم توسعه محصول می‌گرفت. درنتیجه پروژه به‌صورت Real-Time پیاده‌سازی شد و صف کار از آن حذف گردید.

## چالش‌های مربوط به استفاده از پایتون به‌صورت خام

یکی از بزرگ‌ترین بخش‌های پروژه مربوط به پایگاه داده و ارتباط بین جداول مختلف بود. به‌منظور قابل‌گسترش بودن پروژه در آینده از فریمورک جنگو استفاده شد چون هم از ORM بسیار قدرتمندی برخوردار هست و هم دارای Generic Relation بین جداول می‌باشد.

با توجه به اینکه قصد داریم پروژه در آینده گسترش یابد از مفهوم رابطه کلی استفاده می‌کنیم به این صورت که با اضافه شدن انواع مختلف محتوا متنی (پیام کانال، اطلاعات صفحه وب و غیره) و یا محتوا تصویری بتوانیم به‌راحتی ارتباط جدیدی بین جداول داشته باشیم تا نیازی به پیاده‌سازی مفاهیم جدید و پیچیده‌ای نباشد. این امر باعث گسترش‌پذیری و آسودگی برنامه پیاده‌سازی شده می‌شود. علاوه بر موارد ذکرشده فریمورک جنگو توابع و ماژول‌های بسیار زیادی دارد که کار پیاده‌سازی را بسیار آسان‌تر می‌کند و باعث می‌شود تیم توسعه انرژي خود را برای مباحث پیچیده‌تر بگذارند.

## چالش‌های مربوط به ربات تلگرام

در این بخش ابتدا به‌منظور پیاده‌سازی ربات بررسی‌های لازم انجام شد. نتایج به این صورت بود که با استفاده از زبان پایتون رباتی به‌منظور جمع‌آوری داده‌ها نوشته شود. کتابخانه‌ی python-telegram-bot به‌منظور پیاده‌سازی این ربات استفاده شد. به هنگام پیاده‌سازی طبق بررسی‌های مجدد انجام‌شده، بیان‌شده است که ربات‌های نوشته‌شده در تلگرام نمی‌توانند عضو کانال‌های عمومی تلگرام شوند. درنتیجه به‌منظور برطرف کردن این مشکل به‌جای استفاده از ربات، از یک واسطی به نام Telethon که در زبان پایتون پیاده‌سازی شده است استفاده شد.

کتابخانه‌ی Telethon مستقیماً به اکانت تلگرامی کاربران متصل می‌شود و سپس با استفاده از جست‌وجو و پیدا کردن کانال می‌توان بر روی بخش‌های مختلف پیام‌ها پیمایش کند و داده‌ها را استخراج کند. این کتابخانه توابع بسیار زیادی را در اختیار توسعه‌دهندگان قرار می‌دهد که پیاده‌سازی را ساده می‌کند.

### مهیاسازی و آماده‌سازی‌های اولیه

در ابتدا نیاز است که اکانتی برای ربات درست شود. سپس کتابخانه‌ی Telethon با استفاده از رمز مربوط به نرم‌افزار تحت وب به اکانت مربوطه متصل می‌شود و اعتبار سنجی انجام می‌دهد. این اعتبار سنجی به‌منظور امنیت بیشتر دومرحله‌ای بوده. بعد از اعتبار سنجی می‌توانید مانند هر کاربری کانال‌های تلگرامی را پیداکرده و از محتوای آنان استفاده کنید.

البته با توجه به تحریم‌ها و فیلترینگ هوشمند در کشور نیاز به یک فیلترشکن داریم تا بدون مشکل متصل شویم. درصورتی‌که فیلترشکن موجود نباشد بعد از چند ثانیه تمامی بسته‌های ارسال‌شده Timeout می‌شوند و ارتباط بسته می‌شود.

### پیاده‌سازی ربات تلگرامی با استفاده از Telethon

در ابتدا با استفاده از app\_id و hash مربوطه که از سایت تلگرام دریافت شده است به اکانت ربات با استفاده از دستور connect() متصل می‌شویم. توجه کنید بیش از یک اتصال می‌تواند برقرار باشد و برای هر اتصال نشست جداگانه ایجاد می‌شود.

در مرحله بعدی لیست کانال‌های انتخاب‌شده را از پایگاه داده می‌خواند و با پیمایش بر روی این لیست کانالی را که اطلاعات از آن جمع‌آوری می‌شود، قابل‌دسترس است. بعد از دسترسی به کانال‌های تلگرامی شاخص‌های منتخب پیام‌ها استخراج‌شده و در پایگاه داده وارد می‌شوند. این استخراج تنها پیام‌های محدودی را برمی‌دارد و این محدودیت توسط توسعه‌دهنده مشخص می‌شود. شاخص‌های انتخاب‌شده برای پیام‌ها به‌صورت زیر می‌باشند:

* ID: شماره‌ی اختصاص‌یافته‌ی یکتا به پیغام در کانال
* Text: متن پیغام نوشته‌شده
* Pub\_date: تاریخ دقیق پیغام منتشرشده
* Forwards: تعداد forward های این پیغام در این متغیر ذخیره می‌شود.
* Name: نام کانالی که پیام از آن دریافت شده است.
* From\_id: شماره آی دی کاربری که این پیام را ارسال کرده است.
* Forward\_from: شماره آی دی کاربری که این پیام از آن ارسال‌شده است.
* Edit\_date: در صورت ویرایش، تاریخ ویرایش پیام ذخیره می‌شود.
* Edit\_hide‍: درصورتی‌که ادمین کانال تغییری را اعمال کرده باشد و با استفاده از گزینه‌های تلگرام، پیغام edited را از پست حذف کرده باشد، مقدار آن True خواهد شد.
* Is\_reply: آیا پیغام مربوطه reply به پیغام دیگری می‌باشد یا خیر.
* Num\_replies: تعداد پیام‌هایی که به این پیام reply شده‌اند.
* Reply\_to: شماره آی دی پیامی که این پیام به آن reply کرده است.
* Views: بیانگر تعداد دفعاتی که در لحظه دریافت پیام توسط بات، دیده‌شده است.

برنامه به‌صورت کلی از ساعت 8 صبح روزهای شنبه تا پنج‌شنبه به‌غیراز ایام تعطیل تا ساعت 1 اجرا می‌شود. طبق بررسی‌ها و پایش های کانال‌های استفاده‌شده به نظر می‌رسد کانال‌های بورسی در بین این ساعات بیشترین فعالیت خود را دارند.

روند اجرای ربات بدین‌صورت می‌باشد که در ابتدا تابعی برای استفاده‌ سرور از آن نوشته‌شده است. این تابع سه ورودی می‌گیرد که به ترتیب ابتدا نام کانال، تعداد پیام‌ها و تاریخ پیام‌ها را دریافت می‌کند. سپس به دنبال پیام‌هایی از یک کانال می‌رود که ابتدا تاریخ انتشارشان از تاریخی ورودی داده‌شده است بزرگ‌تر باشد و سپس با استفاده از یک حلقه‌ for به تعداد پیام‌های ورودی داده‌شده عملیات دریافت و ذخیره‌سازی پیام‌ها را انجام می‌دهد. بعد از دریافت پیام‌ها آن‌ها را در پایگاه داده ذخیره می‌کنیم.

### چالش‌های پیاده‌سازی ربات تلگرامی

در ابتدا به‌منظور استفاده از ربات و دسترسی داشتن به تلگرام به‌وسیله‌ی پایتون، نیاز است تا در خود سایت تلگرام کاربری که می‌خواهد به تلگرام متصل شود، فرآیند تصدیق را طی کند. بعد از طی کردن این مراحل، سایت تلگرام کدی را تحت عنوان api\_id و api\_hash نمایش می‌دهد که کتابخانه‌ی Telethon از این دو رمز برای ورود به تلگرام استفاده می‌کند. فرآیند طی شده نیاز به فیلترشکن دارد. همچنین در هنگام اجرای کد نیز فیلترشکن باید متصل باشد. (دقت شود استفاده از پروکسی در اجرای کد بررسی شد، اما در هنگام اجرای اولیه کد خطا داده شد)

چالش بعدی در ذخیره‌سازی و مقایسه‌ی زمان‌ها و فیلد Date در پیام‌های خود تلگرام می‌باشد. به دلیل متفاوت بودن time zone مقایسه به‌راحتی انجام نمی‌شود. ابتدا باید time zoneهای پیام‌ها را با تاریخ مطرح‌شده یکسان‌سازی کرد و سپس عملیات مقایسه و ذخیره‌سازی را انجام داد.

چالش مهم بعدی هنگام ذخیره‌سازی پیام‌ها مشاهده شد. مشکل بدین‌صورت بود که در ابتدا پایگاه داده با استفاده از SQL به‌صورت کدهای پایتونی خام، پیاده‌سازی شده بود که به هنگام ذخیره‌سازی داده‌هایی که در آن‌ها حروف ‘(‘ و ‘)’ وجود دارد، خطابه وجود می‌آید. (این خطا ضمنی بوده و مشکل امنیتی تزریق کد را به همراه خواهد داشت) به همین جهت پیاده‌سازی دیتابیس با استفاده از جنگو ابتدا بررسی و سپس اجرا شد.

## چالش‌های پیاده‌سازی پایگاه داده

در بخش پیاده‌سازی پایگاه داده با توجه به اینکه متن‌هایی که از کانال‌های موردنظر به دست می‌آید همگی خام هستند و نیاز به پردازش یا عملیات Text Filtering دارند، به نظر می‌رسد که با استفاده از MongoDB در ابتدا بتوان داده‌ها را ذخیره کرد و سپس با استفاده از الگوریتم‌های احتمالی مانند استفاده از دیکشنری، بتوان داده‌ها را از پیام‌ها خام استخراج کرد. ولی با توجه به توضیحات داده‌شده در بخش‌های نخستین این فرایند بررسی و رد شد. همچنین به منظور ایجاد ارتباط و ذخیره سازی داده‌ها، یک API نوشته می‌شود تا ربات تلگرامی استفاده شده بتواند این داده ها را نیز به سمت پایگاه‌داده انتقال دهد.

## چالش‌های مربوط به تحلیلگر متن

بعد از ذخیره‌سازی داده‌ها در داخل پایگاه‌داده نیاز است تا هر کدام از پیام های ذخیره‌شده تحلیل شوند. تحلیل پیام‌ها بدین صورت می‌باشد که تعیین می‌کند آیا پیام مورد نظر سیگنال می‌باشد یا خیر و در صورتی که سیگنال بود بررسی می‌کند آیا سیگنال فروش است یا خرید؟

### پیاده‌سازی تشخیص سیگنال

به منظور تشخیص سیگنال بودن یا نبود الگوریتمی مبتنی بر شمارش کلمات پیاده‌سازی شد. در ابتدا یک فایلی از تمامی کلید واژه‌های ممکن برای تشخیص سیگنال بودن یک پیام یا نبود آن تهیه شد. به‌منظور تهیه‌ی این لیست، پیام‌های کانال‌های تلگرامی به‌صورت دستی بررسی شدند تا کلمات مشترک آن‌ها بدست آید. کلماتی مانند خرید مهم، سیگنال خرید، سیگنال فروش به‌عنوان مثالی از این کلمات کلیدی می‌باشند.

کلمات کلیدی پیداشده به صورت دستی نیز دسته‌بندی شدند. بیش از ۶۵۰ کلمات کلیدی پیداشده و دسته‌بندی شدند. پنج دسته‌ مربوط به هر کلیدواژه به‌صورت زیر می‌باشند.

* غیرقطعی
* خرید
* فروش
* گروه سهام
* اخبار

برخی از کلمات کلیدی مربوط به خریدوفروش می‌باشند و برخی از کلمات نیز خرید یا فروش بودن سیگنال را تعیین نمی‌کنند اما می‌توانند سیگنال باشند. درنتیجه یک دسته‌ای مربوط به غیرقطعی در نظر گرفته شد. همچنین دسته‌ای تحت عنوان گروه اخبار در نظر گرفته شد چراکه برخی از کانال‌های تلگرامی به برخی از اخبار در مورد سیاست و اجرای توافق‌نامه‌ها و سیاست بین‌الملل مانند برجام، می‌پرداختند که تحلیل آن‌ها به‌احتمال‌زیاد سیگنالی برای بورس نخواهند بود.

در پایان کار به‌منظور افزایش دقت و کارایی الگوریتم به هرکدام از کلمات کلیدی ارزشی بین صفرتا صد نسبت داده شد. ارزش نسبت داده‌شده به هرکدام از کلمات به این دلیل نسبت داده شد که بتوان بین کلمات یک دسته نیز مقایسه انجام داد، چراکه برخی از کلمات در دقت و اطمینان حاصل کردن نسبت به سیگنال بودن یا نبودن تفاوت داشتند. به‌عنوان‌مثال کلمه‌ای مانند سیگنال خرید در مقابل نگهداری دارایی ارزش خیلی بالاتری دارد و در تعیین نوع سیگنال یا سیگنال بودن یا نبودن کمک می‌کند.

ارزش‌گذاری نیز به‌صورت کاملاً دستی و با بررسی‌های کانال‌های تلگرامی صورت گرفته است. لازم به ذکر است این ارزش‌گذاری‌ها بررسی‌شده‌اند و در داده‌های مختلف تست‌شده‌اند تا بیشترین دقت ممکن را داشته باشند. ادامه‌ الگوریتم بدین‌صورت می‌باشد که متن یک پیام را دریافت می‌کند و در ابتدا یک لیستی از ویژگی‌ها داخل یک دیکشنری درست می‌کند. ویژگی‌ها به‌صورت زیر هستند.

* + Is\_signal: آیا پیام سیگنال می‌باشد یا خیر.
  + Is\_buy: آیا سیگنال خرید است یا خیر.
  + Is\_sell: آیا سیگنال فروش است یا خیر.
  + Category: لیستی از نوع صنعتی که این پیام به آن اشاره می‌کند. (مانند دارویی، خودرو)

بعدازاین مرحله حلقه‌ی for بر روی لیست کلیدواژه‌های سیگنال بودن پیمایش می‌کند و تعداد کل کلماتی که به‌عنوان کلیدواژه بوده‌اند و در داخل متن رؤیت شده‌اند را محاسبه می‌کند. همچنین با استفاده از یک حلقه‌ی for دیگر، کلماتی که در دسته‌بندی اخبار بوده‌اند را نیز می‌شمارد.

برای تشخیص سیگنال بودن دو عدد بدست آمده در مرحله‌ی قبل را از یک دیگر کم می‌کند و برروی تعداد کلمات کل متن تقسیم می‌کند تا عملیات نرمال‌سازی و مقایسه‌ی آن بهتر صورت پذیرد. اگر عدد بدست آمده از حاصل عملیات بزرگ‌تر از ۰.۴۴۳ باشد، پیام به‌عنوان سیگنال در نظر گرفته می‌شود. عدد ۰.۴۴۳ عددی تصادفی نیست و با بررسی‌های گوناگون به‌عنوان بهترین مقدار انتخاب‌شده است. این مقدار می‌تواند پویا باشد ولی پیاده‌سازی آن از محدوده پروژه خارج است. تشخیص سیگنال خرید یا فروش بودن نیز به همین صورت پیاده‌سازی انجام‌شده است.

علت این کاهش مقادیر این است که گاهی پیام‌ها سیگنال نمی‌باشند و یا پیام‌های طولانی می‌باشند که حاوی چندین اطلاعات است و پایش را دشوار می‌کند. با کاهش مقادیر سیگنال از اخبار احتمال خطا را کاهش می‌دهیم.

در ارتباط با ستون هایی که سیگنالشان true هستش اما دو ستون is\_sell و is\_buy هردو false هستند، به این دلیل می‌باشد که الگوریتم پیاده سازی شده دارای یک سری محدودیت ها می‌باشد.

تمامی کلمات هر سیگنال‌دهی دسته‌بندی شده‌اند و به هرکدام از آن ها ارزشی داده شده است با توجه به کلمه.

به عنوان مثال کلمه‌ی "خرید" در مقایسه با کلمه‌ی "جامانده" ارزش بالاتری دارد که این ارزش گذاری در چند مرحله آزمون و خطا بدست آمده است.

سپس با استفاده از فرمولی، هر پیام نرماله‌سازی شده است و مقداری برای threshold در نظر گرفته شده. در صورتی که از آن مقدار بزرگتر شود، تشخیص داده خواهد شد که سیگنال می‌باشد. فرمول آن در چند پاراگراف قبلی گزارش گزارش شده است.

حال برای سیگنال خرید یا فروش، الگوریتم دارای برخی محدودیت‌ها می‌باشد، از همین الگوریتم بالا برای پیاده‌سازی سیگنال خرید و فروش نیز استفاده شده است. در صورتی که خروجی الگوریتم نامتعارف باشد به عنوان مثال الگوریتم پیام را سیگنال تشخیص نداده باشد اما سیگنال خرید true شود باید پاکسازی انجام بگیرد و دو ستون خرید و فروش برای آن پیام false شوند

حال درصورتی که اطمینان کامل برای تشخیص نوع سیگنال از منظر خرید یا فروش وجود نداشته باشد، هر دو ستون مربوط به خرید و فروش برابر با false قرار داده می‌شوند اما ستون signal برابر با true خواهد شد. در این حالت با توجه به محدودیت های الگوریتم، با توجه به عدم اطمینان، الگوریتم هیچگونه اظهار نظری در ارتباط با نوع سیگنال انجام نخواهد داد و صرفا بیان می‌کند که پیام سیگنال می‌باشد.

### بهبود و افزایش کارایی الگوریتم

به‌منظور بهبود الگوریتم، یک عملیات پاک‌سازی نیز بر روی‌داده‌های بدست آمده انجام شد. عملیات پاک‌سازی بدین‌صورت است که اگر الگوریتم تشخیص داد که پیام سیگنال نیست اما یکی از دو ویژگی is\_buy یا is\_sell مقدارشان True باشد بدین معناست که الگوریتم درست‌کار نکرده است. به همین جهت در صورت False بودن ویژگی is\_signal، حتماً باید دو ویژگی دیگر مقدار False را به خود بگیرند.

همچنین دو ویژگی is\_sell و is\_buy نمی‌توانند هم‌زمان برای یک پیام True باشند. به همین منظور بررسی می‌شود در صورت رخ دادن چنین موردی الگوریتم نمی‌تواند به‌صورت دقیق نوع سیگنال را تعیین کند. درنتیجه مقدار هر دو ویژگی را False قرار می‌دهیم و تنها مقدار is\_signal را True قرار می‌دهیم.

### چالش‌های پیاده‌سازی تحلیلگر پیام

در ابتدا استفاده از الگوریتم‌های هوش مصنوعی نظیر الگوریتم‌های NLP و تحلیل عواطف بررسی شدند اما با توجه به زمان محدود پروژه و همچنین عدم دانش کافی در این حوزه و همچنین تعیین شدن محدوده‌ی پروژه، الگوریتم تحلیلگر به‌صورت دستی و بدون استفاده از کتابخانه‌های هوش مصنوعی پیاده‌سازی شد.

به هنگام پیاده‌سازی الگوریتم فرآیند نرماله کردن از اهمیت بالایی برخوردار است چرا که در بررسی یک متن تعداد کل کلمات می‌تواند تأثیرگذار باشد. به همین دلیل زمانی برای بررسی متن‌های هر پیغام قرار داده شد تا بتوان به تحلیل درست و دقیقی از متن های پیام ها دست پیدا کرد. بدون استفاده از نرماله سازی پیام‌ها بر روی تعداد کلمات پیام امکان دارد تا در تشخیص سیگنال بودن یک پیام یا نبودن آن دچار اشتباه شویم. با استفاده از نرماله‌ سازی، عملکرد الگوریتم به‌صورت بهینه‌تری کار خواهد کرد.

# فصل چهارم چالش‌های تیم تحلیل چالش‌های مربوط به نرم‌افزار Power BI

Power BI یک محصول نرم‌افزاری تجسم داده تعاملی است که توسط مایکروسافت با تمرکز اصلی بر هوش تجاری توسعه‌یافته است. این بخشی از پلتفرم مایکروسافت پاور است. Power BI مجموعه‌ای از خدمات نرم‌افزاری، برنامه‌ها و رابط‌ها است که باهم کار می‌کنند تا منابع نامرتبط داده را به بینش‌های منسجم، غوطه‌ور کننده بصری و تعاملی تبدیل کنند. داده‌ها ممکن است با خواندن مستقیم از پایگاه داده، صفحه وب یا فایل‌های ساختاریافته مانند صفحات گسترده، CSV، XML و JSON وارد شوند.

برای نمایش خروجی بدست آمده و تحلیل داده‌ها از نرم‌افزار Power BI استفاده می‌کنیم. دلیل این انتخاب با بررسی‌های صورت گرفته این است که نرم‌افزاری بسیار گسترده است و به‌آسانی می‌توان خروجی تیم توسعه نرم‌افزار را در قالب فایل اکسل، SQL گرفته و تحلیل‌های بسیاری را انجام دهیم. یکی از نقاط قوت این نرم‌افزار قابلیت اجرا کردن دستورات SQL برای فیلتر کردن روی‌داده‌ها می‌باشد.

## چالش‌های سخت‌افزاری مربوط به تیم تحلیل

یکی از چالش‌های که با آن مواجه شدیم نبودن نرم‌افزار Power BI برای کامپیوتر‌های اپل بود. اپل در سال ۲۰۱۹ نسل جدید پردازنده‌های M1 خود را رونمایی کرد. این پردازنده‌ها از خانواده پردازنده‌های ARM می‌باشند که معماری متفاوتی نسبت به پردازنده‌های ۳۲ و ۶۴ بیتی دارند. ازآن‌جهت که عده کمی (کمتر از ۱۰ درصد کاربران) از این کامپیوتر‌ها استفاده می‌کنند خیلی از نرم‌افزار‌ها از معماری ARM پشتیبانی نمی‌کنند. به همین دلیل مجبور شدیم تنها از دو سیستم دیگر استفاده کنیم که به علت کند‌تر بودن و قدیمی‌تر بودن چالش‌هایی را به همراه داشته‌اند.

چالش بعدی که با آن مواجه شدیم خراب شدن یکی از سیستم‌های اعضا تیم تحلیل است. این امر به‌شدت سرعت ما‌ را پایین آورد ولی بعد از بررسی‌های صورت گرفته لپتاپی از نزدیکان قرض کردیم و از نسخه وب نرم‌افزار استفاده کردیم. این چالش باعث شد مقداری از برنامه پیاده‌سازی و توسعه عقب بیفتیم ولی با مقداری کوشش به برنامه رسیدیم و ادامه دادیم.

## چالش‌های مربوط به ایجاد حساب

برای استفاده از این نرم‌افزار می‌بایست حساب کاربری خود را داشته باشیم که با توجه به تحریم‌های موجود کاری دشوار می‌باشد. درصورتی‌که حساب کاربری نداشته باشیم امکان استفاده از ابزار‌های نرم‌افزار وجود ندارد.

برای ایجاد حساب کاربری ابتدا از ایمیل دانشگاه استفاده کردیم تا از امکانات نرم‌افزار بهره ببریم. در مرحله بعدی با وارد شدن به‌حساب توانستیم از امکانات استفاده کنیم و نمودار‌های دلخواه را رسم کنیم.

## چالش‌های مربوط به فرمت ورودی Power BI

با توجه به اینکه از پایگاه داده SQL در پیاده‌سازی استفاده کردیم می‌توانستیم خروجی پایگاه داده را به‌سادگی فیلتر کنیم و در هر فرمتی خروجی دهیم. همچنین می‌توانستیم فایل database.sqlite3 موجود در پیاده‌سازی را در Power BI به‌صورت مستقیم وارد کنیم و با سینتکس SQL فیلتر‌های لازم را اعمال کنیم.

در تولید نمودار‌ها هر دو روش استفاده شده به صورتی که برای تعدادی از نمودار‌ها از فایل اکسل و برای تعدادی دیگر از SQL استفاده کردیم. در استفاده از SQL به سادگی می‌توانستیم فیلتر‌هایی را اعمال و تقریباً اتومات نمودار‌ها را تشکیل دهیم. بنابر‌این برای پروژه‌های بعدی تنها از SQL استفاده خواهیم کرد.

## آموزش و یادگیری Power BI

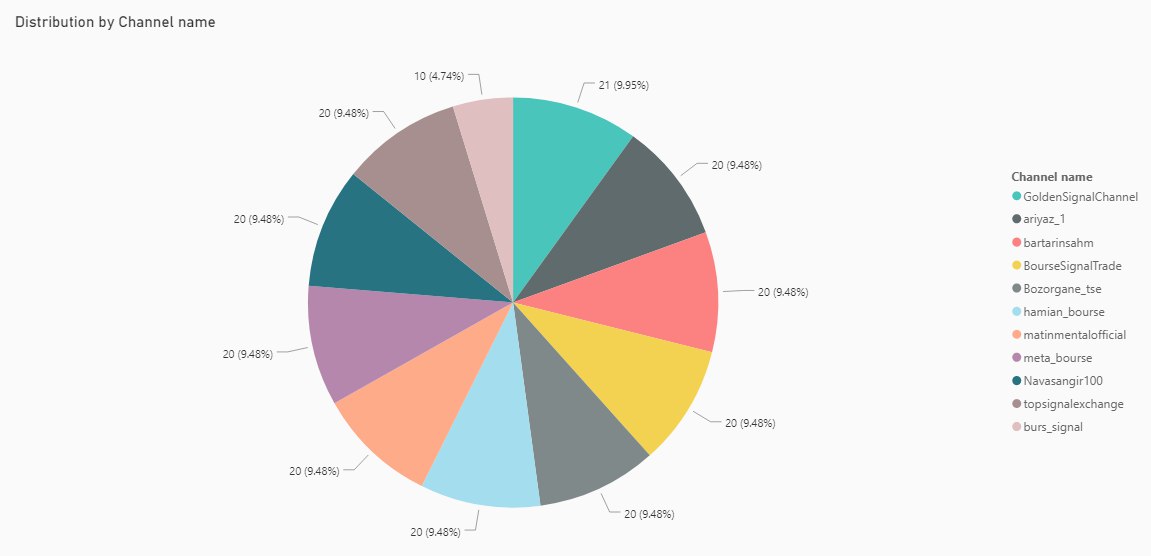
به‌منظور آموزش و یادگیری این نرم‌افزار از فضا‌های مجازی مانند یوتیوب استفاده کردیم. همچنین مستندات ماکروسافت بسیار کامل و دقیق بودند و کمک بسیاری به یادگیری کردند. همچنین فرمی برای پرسش و پاسخ‌های فراوان وجود داشت تا درصورتی‌که به مشکلی شناخته‌شده برخوردیم به‌سادگی مشکل را برطرف کرده و بر آن غلبه کنیم.

### یادگیری با داده آموزشی

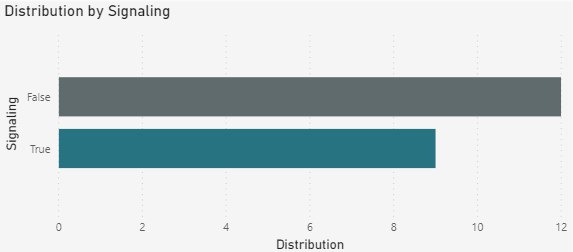
تیم تحلیل به‌موازات تیم توسعه به یادگیری نرم‌افزار پرداختند. این یادگیری و ارتقا دانش زمانی حدود ۱۲ ساعت در هفته طول می‌کشید و این روند به‌اندازه یک اسپرینت و نصف طول کشید. در انتها تیم تحلیل به تسلط نسبی و خوبی رسیدند و آمادگی این را داشتند تا با داده تصادفی ولی در فرمت خروجی نرم‌افزار توسعه‌یافته کار کنند.

تیم توسعه فایل اکسلی که حاوی تمامی ردیف‌های جدول خروجی پایگاه داده هست را در اختیار تیم تحلیل قراردادند. تیم تحلیل نیست با کمک این داده آموزشی نمودار‌های خام و اولیه‌ای تولید کردند. درنهایت در جلساتی تصمیم گرفتیم از چه داده‌هایی نمودار تولید کنیم و نمودار‌های تولیدی به چه شکلی باشند.

به‌عنوان‌مثال از نمودار دایره‌ای برای نمایش تعداد پیام استخراج‌شده از کانال‌های مختلف را با یکدیگر مقایسه می‌کنیم (عکس ۴-۱)، و یا با استفاده از نمودار افقی تعداد پیام‌هایی که سیگنال هستند یا نیستند را در کانال مربوطه با یکدیگر مقایسه کنیم (عکس ۴-۲).



عکس ۴-۱



عکس ۴-۲

## امکانات Power BI

برخی از امکانات بسیار کاربردی نرم‌افزار که مورد استفاده قرار گرفته‌اند لیست و توضیح داده شده‌اند.

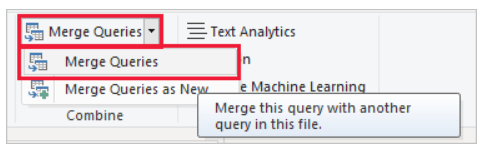
## ترکیب کردن کوئری‌ها

اولین این موضوعات ترکیب کردن کوئری‌ها می‌باشد. دو راه اصلی برای ترکیب کوئری‌ها وجود دارد:

* ادغام: وقتی یک یا چند ستون دارید که می‌خواهید به پرس و جو دیگری اضافه کنید، پرس و جوها را ادغام می‌کنید.
* الحاق: هنگامی که ردیف های دیگری از داده دارید که می‌خواهید به یک پرس و جوی موجود اضافه کنید، پرس و جو را اضافه می‌کنید.

در این حالت می‌خواهیم کوئری‌ها را با هم ادغام کنیم. برای انجام آن، این مراحل را دنبال می‌کنیم (عکس ۴-۳).

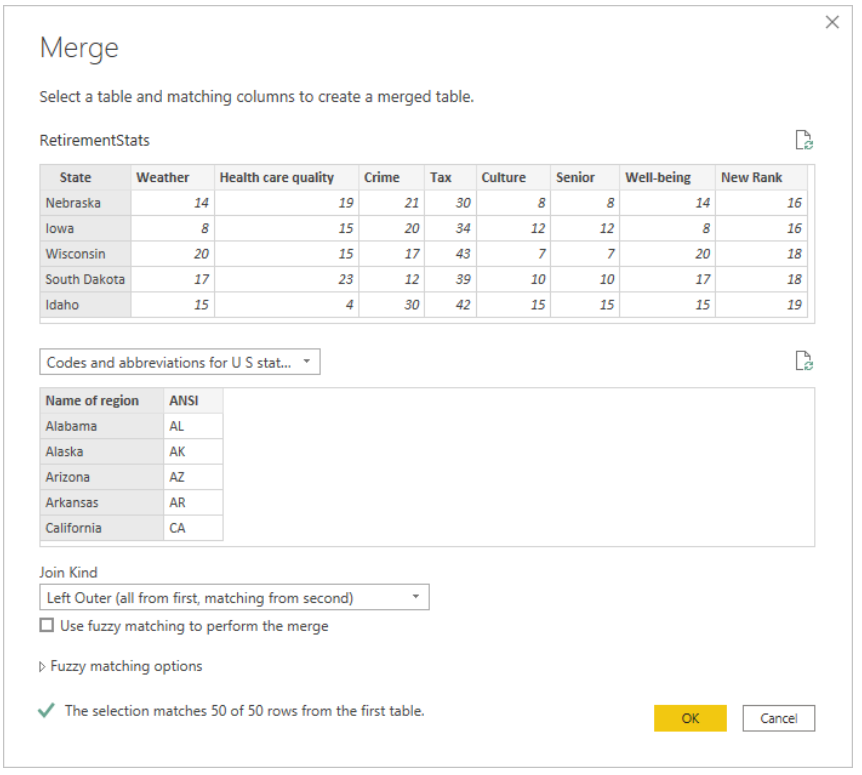
* از پنجره سمت چپ ویرایشگر Power Query، عبارتی را که می‌خواهید پرس و جوی دیگر در آن ادغام شود، انتخاب کنید. در این مورد،RetirementStats است.
* Merge Queries > Merge Queries را از تب Home در نوار انتخاب کنید.



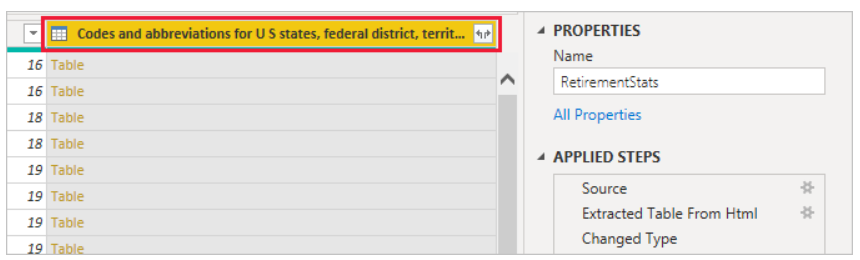
عکس ۴-۳

* از جدول RetirementStats State را انتخاب کنید، سپس عبارت StateCodes را انتخاب کنید (عکس ۴-۴).
* OK را انتخاب کنید.

ویرایشگر Power Query یک ستون جدید در انتهای پرس و جو ایجاد می کند که حاوی محتویات جدول است که با کوئری موجود ادغام شده است. تمام ستون‌های پرس و جو ادغام شده در ستون فشرده می‌شوند، اما می‌توانید جدول را گسترش دهید و ستون‌هایی را که می‌خواهید اضافه کنید (عکس ۴-۵).



عکس ۴-۴



عکس ۴-۵

* برای گسترش جدول ادغام شده، و انتخاب ستون‌هایی که شامل شوند، نماد گسترش هستند را انتخاب کنید.
* در این حالت فقط ستون State Code را می‌خواهیم. ستون را انتخاب کنید، Use original column name as prefix را پاک کنید و سپس OK را انتخاب کنید. اگر چک باکس را برای استفاده از نام ستون اصلی به عنوان پیشوند انتخاب کرده بودیم، ستون ادغام شده NewColumn.State Code نامیده می‌شد. ما اکنون یک پرس و جو (جدول) داریم که دو منبع داده را ترکیب می‌کند، که هر کدام برای رفع نیازهای ما شکل گرفته اند. این پرس و جو می تواند به عنوان پایه ای برای بسیاری از ارتباطات داده‌های اضافی و جالب، باشد.
* برای اعمال تغییرات و بستن Power Query Editor، گزینه Close & Apply را از تب صفحه اصلی انتخاب کنید. مجموعه داده تبدیل شده در Power BI Desktop ظاهر می‌شود و آماده استفاده برای ایجاد گزارش است.

یک عملیات append یک پرس و جو جدید ایجاد می‌کند که شامل تمام ردیف های یک پرس و جو اول و سپس تمام ردیف‌های یک پرس و جو دوم است. عملیات append حداقل به دو پرس و جو نیاز دارد. این پرس و جوها همچنین می‌توانند بر اساس منابع داده‌های خارجی مختلف باشند.

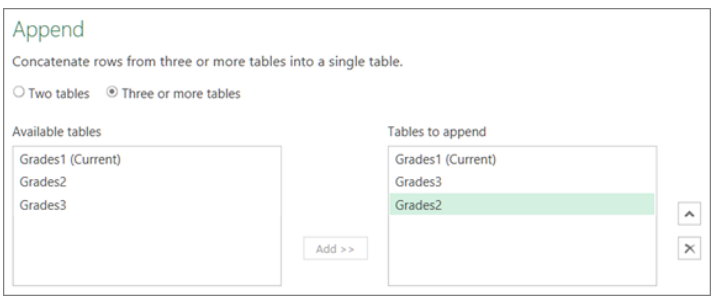
عملیات الحاق بر اساس نام سرصفحه های ستون در هر دو جدول است و نه موقعیت ستون نسبی آنها. جدول نهایی دارای تمام ستون‌های منطبق از همه جداول ضمیمه شده است. اگر جدول‌ها ستون‌های منطبق نداشته باشند، مقادیر null به ستون بی‌همتا اضافه می‌شوند. جداول به ترتیبی که انتخاب شده اند و از جدول اصلی شروع می شود، اضافه می شوند.

شما می توانید دو نوع عملیات append را انجام دهید. با یک ضمیمه درون خطی، تا زمانی که به نتیجه نهایی برسید، داده ها را به پرس و جوی موجود خود اضافه می‌کنید. نتیجه یک مرحله جدید در پایان پرس و جو فعلی است. با یک ضمیمه میانی، برای هر عملیات ضمیمه یک پرسش جدید ایجاد می‌کنید.

مراحل الحاق به شرح ذیل می‌باشد:

* برای باز کردن یک پرس و جو، یکی را که قبلاً از ویرایشگر Power Query بارگیری شده است پیدا کنید، یک سلول را در داده‌ها انتخاب کنید و سپس Query > Edit را انتخاب کنید. برای اطلاعات بیشتر به ایجاد، بارگیری یا ویرایش یک پرس و جو در اکسل مراجعه کنید.
* Home > Append Queries را انتخاب کنید. عمل پیش‌فرض انجام یک ضمیمه درون خطی است. برای انجام یک پیوست میانی، فلش کنار دستور را انتخاب کنید و سپس Append Queries را به عنوان New انتخاب کنید.

کادر محاوره ای Append ظاهر می‌شود (عکس ۴-۶).



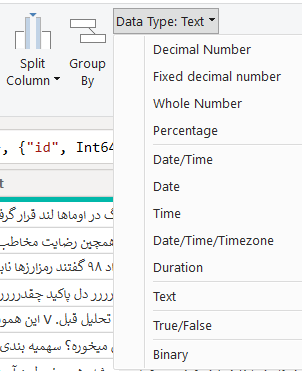
عکس ۴-۶

* تعداد جداولی که می خواهید اضافه کنید را تعیین کنید:
  + دو جدول را انتخاب کنید و سپس جدول دوم را در کادر لیست کشویی برای پیوست انتخاب کنید.
  + سه یا چند جدول را انتخاب کنید. از کادر جداول موجود، جداولی را که می خواهید ضمیمه کنید به جداول اضافه کنید اضافه کنید. از فلش های سمت راست آن کادر برای تغییر ترتیب استفاده کنید.
  + OK را انتخاب کنید.

در نتیجه عملیات فوق اگر در مرحله دوم یک پیوست درون خطی انجام دهید، یک مرحله جدید در پرس و جو فعلی ایجاد می‌شود. می‌توانید به افزودن مراحل به همان جستار برای افزودن عبارت‌های جستجوی اضافی ادامه دهید. اگر در مرحله دوم یک ضمیمه میانی را انتخاب کنید، یک پرس و جو جدید ایجاد می‌شود. می‌توانید به ایجاد پرس و جوهای اضافی ادامه دهید.

#### أنواع مدل داده‌ها (Data types)

در بخش انواع مدل داده‌ها ما می‌توانیم با استفاده در ستون‌های خود از یک نوع داده خاص جهت دسته‌بندی استفاده کنیم همانطور که در تصویر مشاهده می‌کنیم می‌توانیم داده‌ها را به صورت اعداد اعشاری، اعداد صحیح، درصد، روز و زمان و منطقه زمانی، متن، صحیح/غلط و عدد باینری نشان دهیم و با توجه به آن فیلتر شود (عکس ۴-۷).



عکس ۴-۷

Quick measures

بسیاری از محاسبات رایج به عنوان معیارهای سریع در دسترس هستند، که فرمول‌های DAX را بر اساس ورودی های شما در یک پنجره برای شما می‌نویسند. این محاسبات سریع و قدرتمند همچنین برای یادگیری DAX یا تخمین معیارهای سفارشی شما عالی هستند.

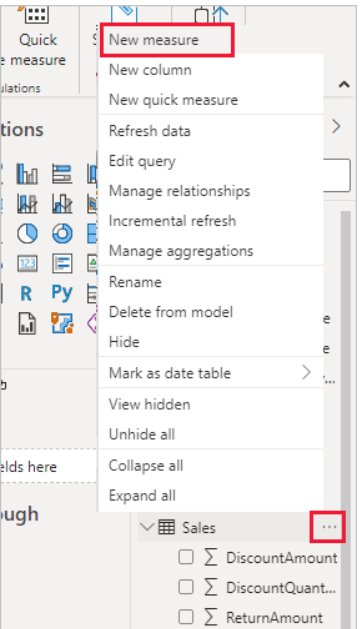
Create a measure

فرض کنید می‌خواهید فروش خالص خود را با کم کردن تخفیف‌ها و بازده‌ها از کل مبالغ فروش تجزیه و تحلیل کنید. برای زمینه ای که در تجسم شما وجود دارد، به معیاری نیاز دارید که مجموع DiscountAmount و ReturnAmount را از مجموع SalesAmount کم کند. هیچ فیلدی برای فروش خالص در لیست فیلدها وجود ندارد، اما شما بلوک‌های ساختمانی برای ایجاد معیار خود برای محاسبه فروش خالص دارید.

برای ایجاد یک اندازه گیری، مراحل زیر را دنبال کنید:

* در قسمت فیلدها، روی جدول فروش کلیک راست کرده یا ماوس را روی جدول نگه دارید و گزینه های بیشتر (...) را انتخاب کنید.
* از منوی ظاهر شده، New measures را انتخاب کنید (عکس ۴-۸)

این اقدام اندازه گیری جدید شما را در جدول فروش ذخیره می کند، جایی که پیدا کردن آن آسان است.

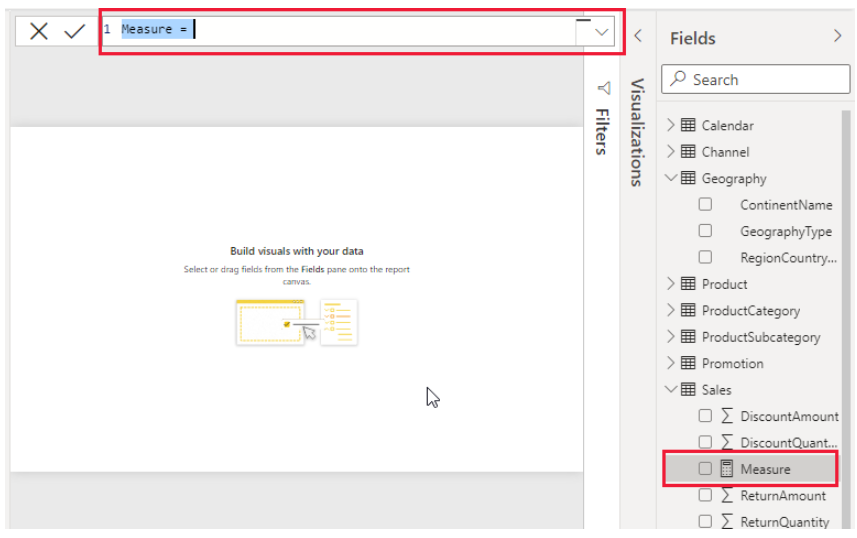


عکس ۴-۸

همچنین می توانید با انتخاب New Measure در گروه Calculations در زبانه Home از روبان Power BI Desktop یک اندازه گیری جدید ایجاد کنید.

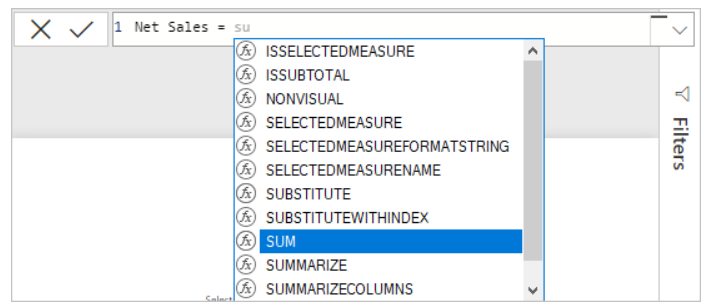
وقتی اندازه ای را از روبان ایجاد می کنید، می توانید آن را در هر یک از جدول های خود ایجاد کنید، اما اگر آن را در جایی که قصد استفاده از آن را دارید ایجاد کنید، پیدا کردن آن آسان تر است. در این حالت ابتدا جدول فروش را انتخاب کنید تا فعال شود و سپس New measure را انتخاب کنید.

نوار فرمول در بالای بوم گزارش ظاهر می شود، جایی که می توانید نام اندازه گیری خود را تغییر دهید و فرمول DAX را وارد کنید (عکس ۴-۹)



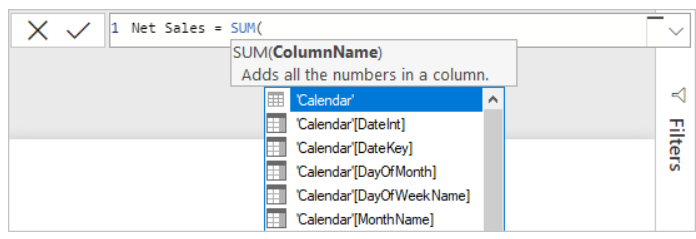
عکس ۴-۹

* به طور پیش فرض، هر اندازه گیری جدید Measure نام دارد. اگر نام آن را تغییر ندهید، اقدامات جدید اضافی به نام‌های Measure 2، Measure 3 و غیره نامیده می‌شوند. از آنجایی که می‌خواهیم این معیار قابل شناسایی‌تر باشد، Measure را در نوار فرمول برجسته کرده و سپس آن را به فروش خالص تغییر دهید.
* وارد کردن فرمول خود را شروع کنید. بعد از علامت مساوی شروع به تایپ Sum کنید. همانطور که تایپ می کنید، یک لیست پیشنهادی کشویی ظاهر می شود که همه عملکردهای DAX را نشان می دهد و با حروفی که تایپ می کنید شروع می شود. در صورت لزوم، برای انتخاب SUM (عکس۴-۱۰) از لیست به پایین بروید و سپس Enter را فشار دهید.



عکس ۴-۱۰

یک پرانتز باز به همراه یک لیست پیشنهادی کشویی از ستون های موجود که می توانید به تابع SUM منتقل کنید ظاهر می شود (عکس ۴-۱۱)

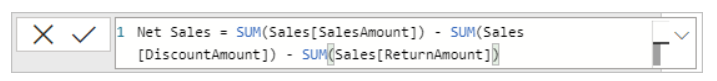


عکس ۴-۱۱

* عبارات همیشه بین پرانتز باز و بسته ظاهر می شوند.

نام ستون قبل از نام جدول، نام کاملا واجد شرایط ستون نامیده می شود. نام ستون‌های کاملا واجد شرایط، خواندن فرمول‌های شما را آسان‌تر می‌کنند.

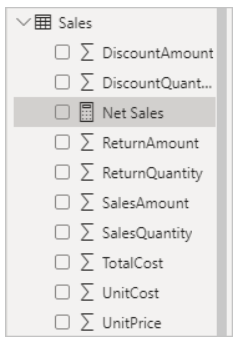
* از لیست انتخاب کنید و سپس یک پرانتز بسته را وارد کنید. خطاهای نحوی اغلب به دلیل گم شدن یا نابجا بودن پرانتز بسته شدن ایجاد می شوند.
* دو ستون دیگر داخل فرمول را کم کنید:
  + بعد از پرانتز بسته شدن عبارت اول، یک فاصله، یک عملگر منهای (-) و سپس یک فاصله دیگر را تایپ کنید.
  + تابع SUM دیگری را وارد کنید و شروع به تایپ DiscountAmount کنید تا زمانی که بتوانید ستون Sales[DiscountAmount] را به عنوان آرگومان انتخاب کنید. یک پرانتز بسته اضافه کنید.
  + یک فاصله، یک عملگر منهای، یک فاصله، یک تابع SUM دیگر با آرگومان Sales[ReturnAmount] و سپس یک پرانتز بسته را تایپ کنید.



عکس ۴-۱۲

* برای تکمیل و اعتبارسنجی فرمول، Enter را فشار دهید یا Commit (نماد علامت تیک) را در نوار فرمول انتخاب کنید.

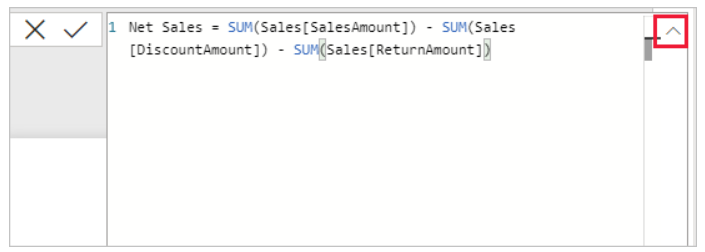
معیار فروش خالص تأیید شده اکنون آماده استفاده در جدول فروش در صفحه فیلدها است (عکس ۴-۱۳).



عکس ۴-۱۳

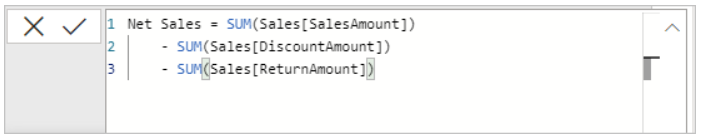
* اگر فضای کافی برای وارد کردن فرمول ندارید یا می‌خواهید آن را در خطوط جداگانه قرار دهید، فلش رو به پایین در سمت راست نوار فرمول را انتخاب کنید تا فضای بیشتری ارائه شود.

فلش رو به پایین به یک فلش رو به بالا تبدیل می شود و یک کادر بزرگ ظاهر می شود (عکس ۴-۱۴).



عکس ۴-۱۴

بخش هایی از فرمول خود را با فشار دادن Alt + Enter برای خطوط جداگانه یا فشار دادن Tab برای افزودن فاصله بین برگه ها جدا کنید (عکس ۴-۱۵).



عکس ۴-۱۵

New column

در این بخش می‌توانیم یک ستون جدید به داده‌های خود اضافه کنیم با استفاده از این بخش داده‌های جدید را به جدول خود اضافه می‌کنیم، در این بخش می‌توانیم نوع داده‌های ورودی به این ستون از جدول و همچنین معیار سنجش این ستون از جدول را مشخص می‌کنیم. در این بخش می‌توانیم بخش‌هایی از داده‌ها را به همدیگر مرتبط می‌کنیم و می‌توانیم داده‌ها را با استفاده از ستون‌ها دسته‌بندی کنیم.

### چالش‌های مربوط به دسته‌بندی مناسب داده‌های خروجی

یکی از چالش‌ها دسته‌بندی مناسب ستون‌های اکسل و SQL می‌باشد. در نرم‌افزار Power BI باید مشخص کنیم کدام ستون‌ها را برای رسم نمودار انتخاب کند. به‌عنوان‌مثال ستونی که اطلاعات زمانی به فرمت سال:ماه:روز ساعت:دقیقه:ثانیه ذخیره می‌کند نباید به‌صورت خام استفاده شود چراکه ممکن است اختلاف‌زمانی ردیف‌ها در حد ساعت و یا روز باشد و نمایش با سال بسیار نمودار نادقیق و اشتباهی تولید می‌کند. برای تولید هر نمودار ستون‌های متفاوتی انتخاب شدند و نتایج مقایسه شدند تا بهترین نمودار انتخاب شود.

### چالش مربوط به اکسل

در اکسل با مشکلی مواجه شدیم. ستون‌هایی که خالی بودند با نام column\_name و داده null مقداردهی شده بودند و این تولید نمودار را با مشکل مواجه کرد به‌طوری‌که مجبور به پاک کردن اطلاعات اکسل و گرفتن خروجی دوباره شدیم. احتمالاً دلیل این امر قدیمی بودن نسخه اکسل بر روی کامپیوتر اعضا بود. بااین‌حال این مشکل درزمانی کوتاه برطرف شد.

### چالش‌های مربوط به رابط کاربری و رنگ‌بندی نمودار‌ها

یکی از چالش‌ها انتخاب نمودار مناسب می‌باشد. درصورتی‌که نمودار اشتباه برای نمایش داده‌هایمان انتخاب کنیم می‌توانیم پیامی نادرست ارسال کنیم. به‌عنوان‌مثال بهتر از پراکندگی را به کمک نمودار Pie-Chart نمایش دهیم و نمودار‌های زمانی را به‌صورت پیوسته نمایش دهیم.

چالش دیگری که با آن روبه‌رو شدیم رنگ‌بندی نمودار‌ها می‌باشد. با توجه به اینکه نمودار‌های گوناگون مفاهیم مختلفی را در بردارند نباید رنگ‌بندی یکسانی نیز داشته باشند در غیر این صورت اطلاعات غلطی را انتشار می‌دهند. سعی کردیم تا برای نمودار‌ها از رنگ‌های گوناگون استفاده کنیم تا بهترین کیفیت را داشته باشند.

### چالش‌های مربوط به مدل‌سازی و ارتباط بین نمودار‌ها

یکی از قابلیت‌های بسیار جذاب نرم‌افزار Power BI توانای برقراری ارتباط بین نمودار‌ها می‌باشد. به این صورت که اگر در نقاط خاصی از نمودار مادر کلیک شود یا تغییری ایجاد شود تمامی نمودار‌های فرزند نیز دچار تغییراتی می‌شوند. در برخی نمودار‌ها از این قابلیت جذاب استفاده کردیم تا اطلاعات را به شکل صحیح و مفید انتقال دهیم.

### چالش‌های مربوط به استفاده از Color Saturation در Tree-map

در تولید برخی نمودار‌ها که قصد داشتیم از Color Saturation استفاده کنیم، به این صورت که نمودار با پیشروی به سبز و با پس‌روی به قرمز متمایل شود و شهود بهتری از پیشروی نمودار در اختیار کاربران قرار می‌دهد.

چالشی که با آن روبه‌رو شدیم به این دلیل بود که بخش Data Color در نسخه‌های جدید Power BI حذف‌شده است و به شکلی ساختار قابل‌استفاده است. اما این ساختار جدید قابلیت‌های نسخه قبلی را ندارد و باعث شد از این نمودار گذر کنیم و به نمودار‌های دیگر بپردازیم.

## چالش‌های مربوط به تولید کلیدواژه‌ها

یکی از رکن‌های اصلی نرم‌افزار تحلیل‌گر استفاده از کلیدواژه‌ها و ارزش‌های آن‌ها بود. در اصل تمام تحلیل و تشخیص سیگنال و عدم سیگنال از همین طریق صورت می‌گرفت. درنتیجه تیم تحلیل در طول اسپرینت اول بیست کانال بورس و سیگنال که کاربران زیادی داشتند را انتخاب کردند و به دنبال کلید‌واژه‌ها گشتند. این کلید‌واژه‌ها به‌صورت یک نگاشت یک به چند بود. به‌عنوان‌مثال کلیدواژه‌های خرید، ارزش، دارای و نگه‌داری سهم معنی سیگنال خرید را می‌دهند.

این کلیدواژه‌ها توسط تیم تحلیل استخراج و در دسترس تیم توسعه قرار گرفت. بعد تیم توسعه این اطلاعات را در پایگاه داده ثبت کردند تا بخش تحلیل‌گر نرم‌افزار از آن‌ها استفاده کند.

## چالش‌های مربوط به ارزش کلید‌واژه‌ها

هر کلیدواژه ارزشی دارد. این ارزش باید از بقیه کلیدواژه‌ها متفاوت باشد تا تحلیل به‌درستی انجام گیرد. به‌عنوان‌مثال ارزش کلیدواژه «خرید» به‌مراتب بالاتر از «شرایط مثبت» می‌باشد. تیم تحلیل در طول اسپرینت اول و دوم کلیدواژه‌های استخراج‌شده را ارزش‌گذاری کردند و این ارزش‌ها را در اختیار تیم توسعه قرار دادند. بعد تیم توسعه این اطلاعات را در پایگاه داده ثبت کردند تا بخش تحلیل‌گر نرم‌افزار از آن‌ها استفاده کند.

## چالش بدست آوردن فرمول تحلیل‌گر

با داشتن کلید‌واژه‌ها و ارزش‌ها می‌توان ارزش خام هر فرایند را بدست آورد. ولی در مواردی مشاهده کردیم پیامی طولانی وجود دارد که از شرایط فعلی بازار و یا سیگنال دهی اشتباه شکایت به عمل می‌آورند و یا پیامی طولانی با چندین سیگنال نامشخص وجود دارد. درنتیجه پیام‌های یک ماه‌ را از کانالی دریافت و ارزش‌گذاری کردیم، این ارزش‌ها را مقایسه کردیم و Threashold نرمال‌ شده‌ای را بدست آوردیم به صورتی که ارزش کلیدواژه‌ها بر ارزش غیر کلیدواژه‌ها باید از مقداری بیشتر باشد در غیر این صورت پیام معتبر نمی‌باشد. می‌توان در تکامل‌های بعدی پروژه روشی برای پیاده‌سازی فرمولی پویا بدست آورد ولی در شرایط فعلی از محدوده پروژه خارج بوده است.

## چالش مقایسه نتایج با سایت کدال

به‌منظور کشف تقلب قصد داریم تا از نتایج تحلیل‌شده توسط نرم‌افزار نمودار‌هایی را تولید کنیم و با استفاده از نمودار‌ها و مقایسه نتایج با اطلاعات سایت کدال کانال‌های مشکوک را شناسایی کنیم و گزارش دهیم. برای این منظور نمودار‌هایی را با کمک اطلاعات ذیل ایجاد می‌کنیم.

* + تنوع پیام‌ها از کانال‌های مختلف.
  + پیک بازدید از هر کانال در ساعات گوناگون روز.
  + تعداد پیام‌هایی که سیگنال هستند/نیستند.
  + تعداد پیام‌هایی که سیگنال خرید/فروش هستند.
  + تعداد دفعات فوروارد شدن پیام کانال‌ها
  + ادیت شدن یا نشدن پیام‌های کانال
  + و ...

در مرحله آخر این نمودار‌ها را با درصد با‌ارزش/کم‌ارزش شدن سهم سیگنال دهی شده مقایسه می‌کنیم. درصورتی‌که باوجود اطلاعات ضدونقیض، شرایط خوب و بد بورس اوراق بهادار و تعداد دفعات سیگنال دهی درست می‌توان تا حد خوبی کشف تقلب انجام دهیم.

چالشی که در این مرحله با آن برخوردیم نیاز به بررسی دستی و بدون استفاده از کامپیوتر بود که کار را بسیار دشوار کرد و درصورتی‌که قصد گسترش پروژه در سطح تمامی شبکه‌های اجتماعی داشته باشیم تیم تحلیل بسیار بزرگ متشکل از متخصصین بسیار زیاد نیاز داریم که هزینه و وقت بسیار زیادی هدر می‌دهد.

ولی در سطح محدوده پروژه همین میزان کافی است و تنها قصد Proof-of-Concept - اثبات مفهوم (POC یا PoC) که به‌عنوان اثبات اصل نیز شناخته می‌شود، تحقق یک روش یا ایده معین به‌منظور اثبات امکان‌پذیری آن، یا اثبات اصولی باهدف تأیید اینکه یک مفهوم یا نظریه دارای پتانسیل عملی است. اثبات مفهوم معمولاً کوچک است و ممکن است کامل باشد یا نباشد - را داشتیم که مشخص کنیم آیا این امر شدنی است یا راه دیگری را باید پیش گرفت.

# فصل پنجم خروجی نرم‌افزار و تیم تحلیل

# خروجی نهایی و نمودار‌های خروجی

به‌منظور تحلیل خروجی و بررسی‌های انجام‌شده از نمودار‌های گوناگون استفاده می‌کنیم. تحلیل‌ها به دو صورت گروهی و تکی صورت می‌گیرد. تحلیل‌های گروهی مربوط به‌تمامی کانال‌ها می‌باشند مانند تعداد بازدید، تعداد بازنشر و غیره درصورتی‌که نمودار‌های تکی داده‌های هر کانال را نسبت به خودش بررسی می‌کند مانند تعداد سیگنال خرید/فروش و یا اوج بازدید کانال‌ها در ساعاتی خاص. در این پروژه کانال‌های ذیل مورد بررسی قرار می‌گیرند. توجه کنید داده‌های استخراج‌شده مربوط به هفته‌های انتهای خرداد ماه سال ۱۴۰۱ می‌باشد.

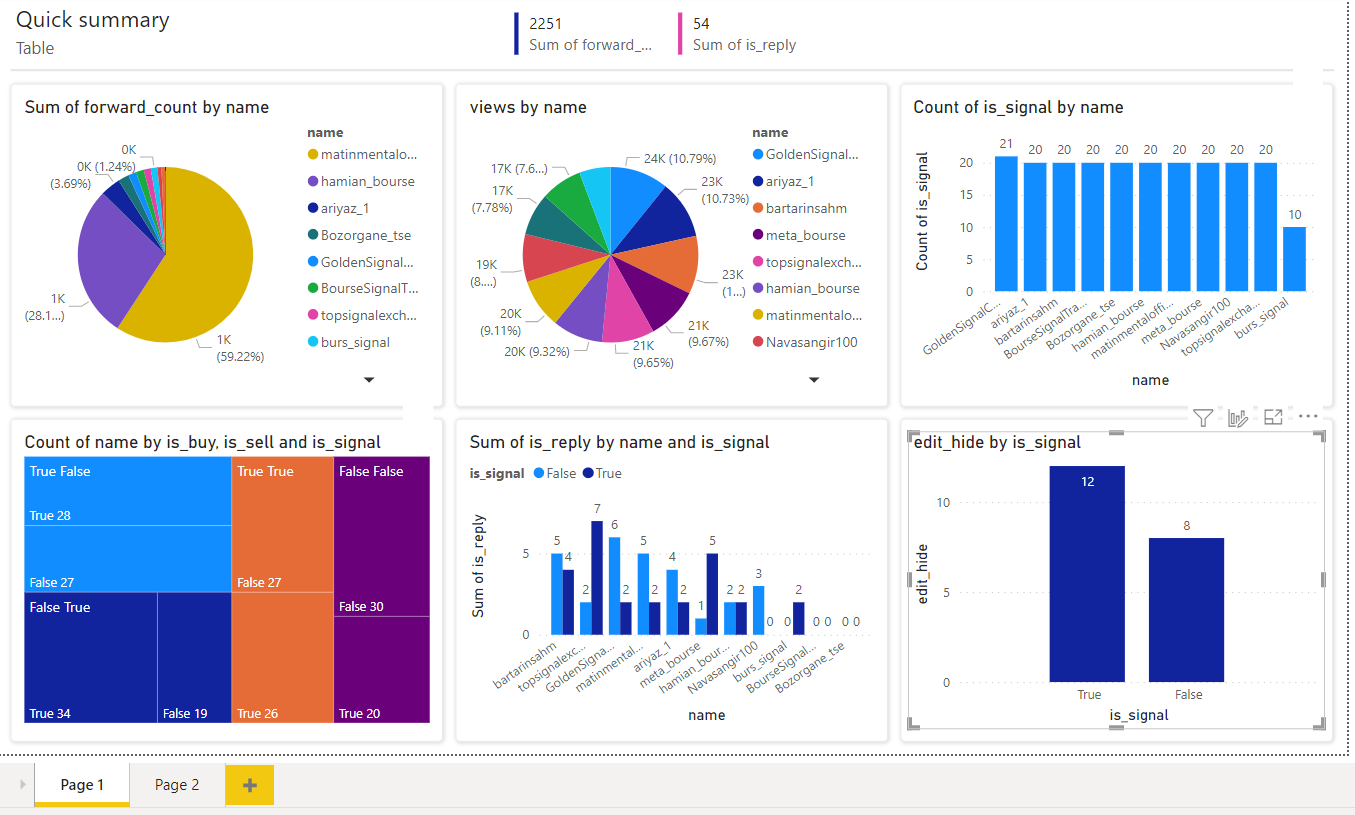
* https://t.me/ariyaz\_1
* https://t.me/bartarinsahm
* https://t.me/topsignalexchange
* https://t.me/Bozorgane\_tse
* https://t.me/BourseSignalTrade
* https://t.me/hamian\_bourse
* https://t.me/GoldenSignalChannel
* https://t.me/Navasangir100
* https://t.me/matinmentalofficial
* https://t.me/burs\_signal
* https://t.me/meta\_bourse

## نمودار‌های گروهی

دو نمودار گروهی تولیدشده که به نمایش و توضیح هریک می‌پردازیم

### نمودار اول: بررسی مختصر

نمودار ۵-۱ نمودار بررسی مختصر کانال‌های انتخاب‌شده می‌باشد. این نمودار از شش نمودار کوچک‌تر تقسیم‌شده که تمامی اجزاء آن به یکدیگر متصل هستند و با تغییر یکی مابقی نیز تغییر می‌کنند. این تغییر در مقادیر و شکل نمودار نمی‌باشد و صرفاً در نمایش اطلاعات می‌باشد.

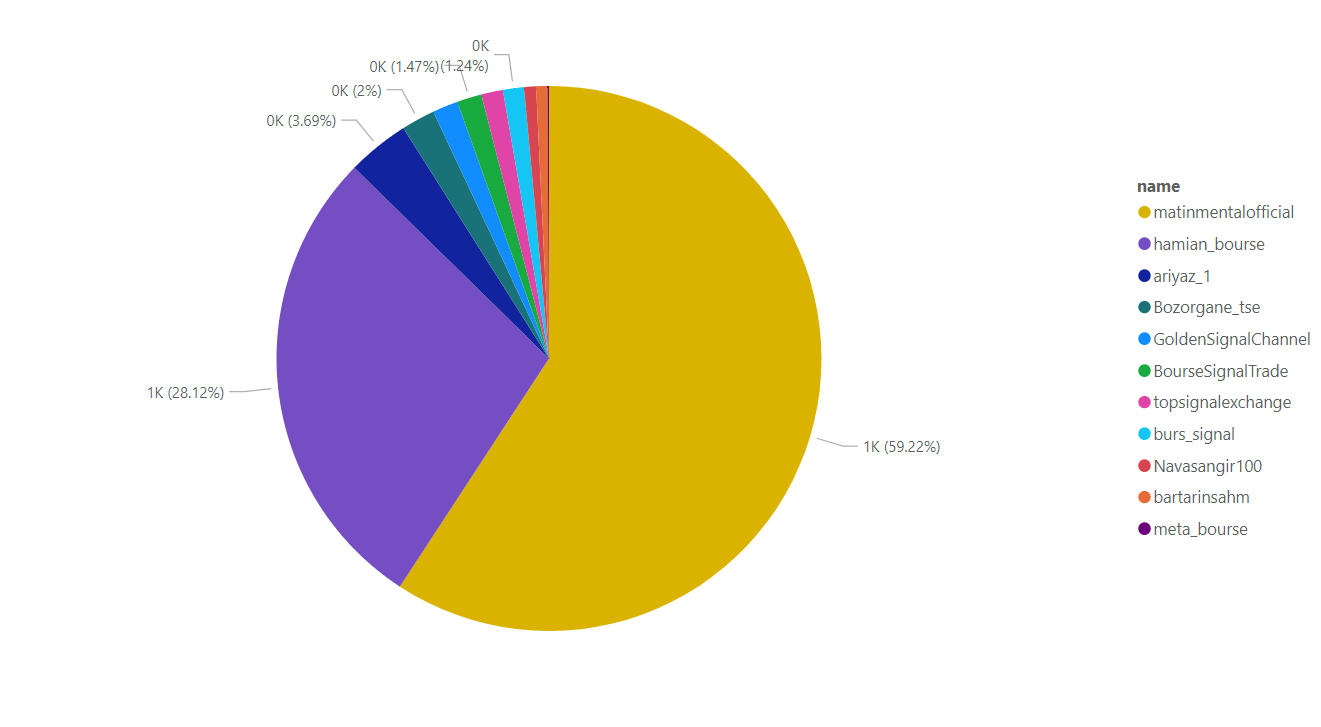


نمودار ۵-۱

یکی از عواملی که کانال متقلب را موردتوجه سازمان بورس اوراق بهادر قرار می‌دهد تعداد بازدید و بازنشر پیام‌هایش است. به این منظور کانال‌های بررسی‌شده را از این منظور با یکدیگر مقایسه می‌کنیم. دو نمودار اول از چپ بالا بیانگر این مفاهیم هستند.

## تعداد بازنشر پیام‌های بارگذاری شده

همان‌طور که گفته شد تعداد بازنشر یکی از فاکتور‌های بسیار مهم در کشف تقلب می‌باشد. درصورتی‌که دو کانال متقلب داشته باشیم اولویت کشف تقلب با کانالی است که بازنشر بیشتری دارد و بر روی بازار بورس اوراق بهادر تأثیر بیشتری می‌گذارد. نمودار ۵-۲ بیانگر این مفهوم می‌باشد.

نمودار ۵-۲

## تعداد بازدید پیام‌های بارگذاری شده

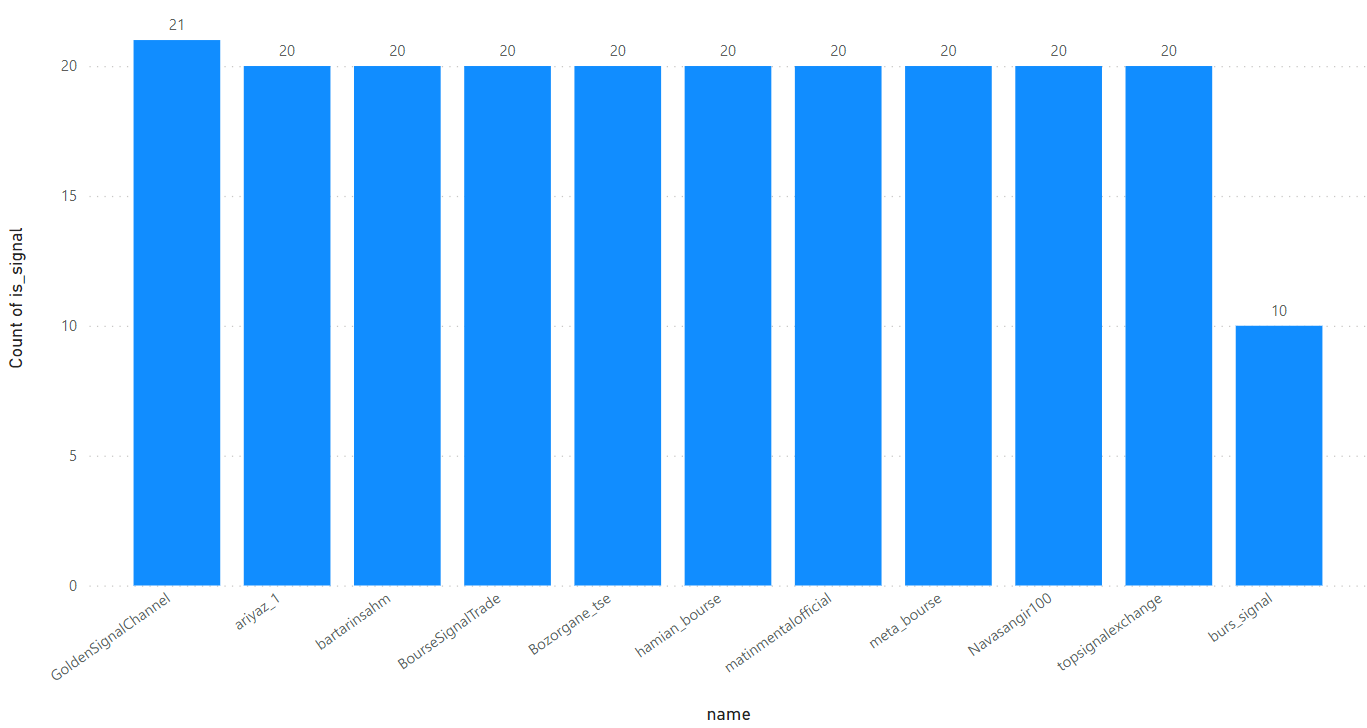
تعداد بازنشر تأثیر مستقیم بر تعداد بازدید پیام‌ها دارد و هرچه تعداد بازنشر بالاتر رود تعداد بازدید پیام‌ها نیز بالاتر می‌رود و بر میزان خطر افزوده می‌شود. نمودار ۵-۳ بیانگر تعداد بازدید کل پیام‌های استخراج‌شده از کانال‌های مربوطه می‌باشد. با مقایسه‌ای ساده به‌درستی گزاره «هرچه تعداد بازنشر بالاتر باشد تعداد بازدید نیز بالاتر می‌باشد» می‌رسیم.

همچنین هرچه تعداد بازدید پیام‌های کانالی بیشتر باشد سیگنال‌های خریدوفروش کانال مربوطه به‌مراتب بیشتر موردتوجه قرار می‌گیرند. پس معیار بعدی که موردمطالعه قرار می‌گیرد تعداد سیگنال‌های داده‌شده توسط هر کانال تلگرامی می‌باشد.

نمودار ۵-۳

## سیگنال‌ دهی انجام شده توسط کانال‌ها

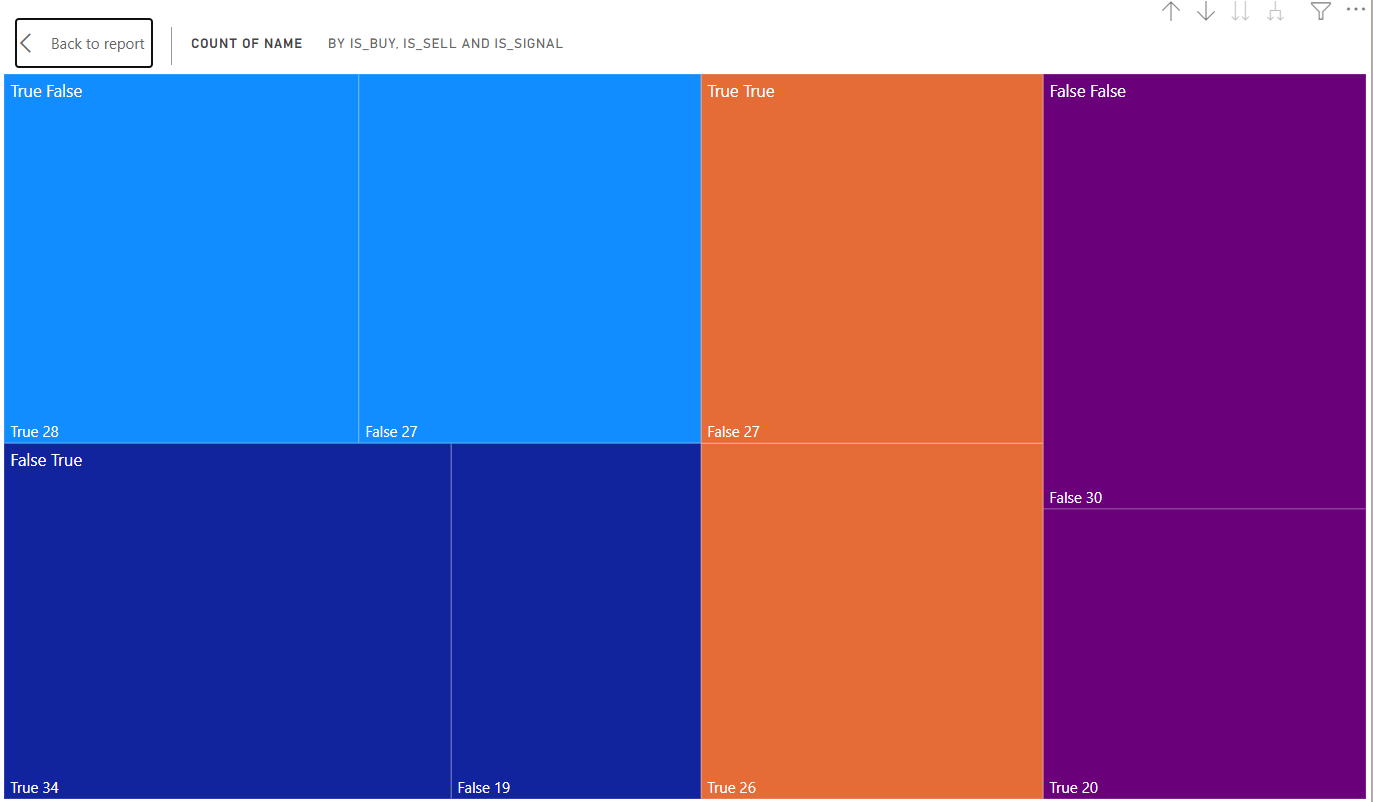
این موضوع که هر کانال چه تعداد سیگنال دهی داشته‌اند در کنار تعداد بازدید و بازنشر عمر بسیار مهمی در کشف تقلب می‌باشد. نمودار ۵-۴ نموداری ستونی می‌باشد که این مقایسه را برایمان آسان می‌سازد.



نمودار ۵-۴

## تعداد سیگنال‌ها، سیگنال‌های خرید و سیگنال‌های فروش

در نمودار ۵-۵ می‌توانید گستردگی سیگنال‌ها، سیگنال‌های خریدوفروش را به‌صورت جمعی مشاهده کنیم. این نمودار به ما کمک می‌کند تا بررسی کنیم کدام کانال‌ها بیشتر در چه سیگنال دهی‌هایی دخیل هستند.

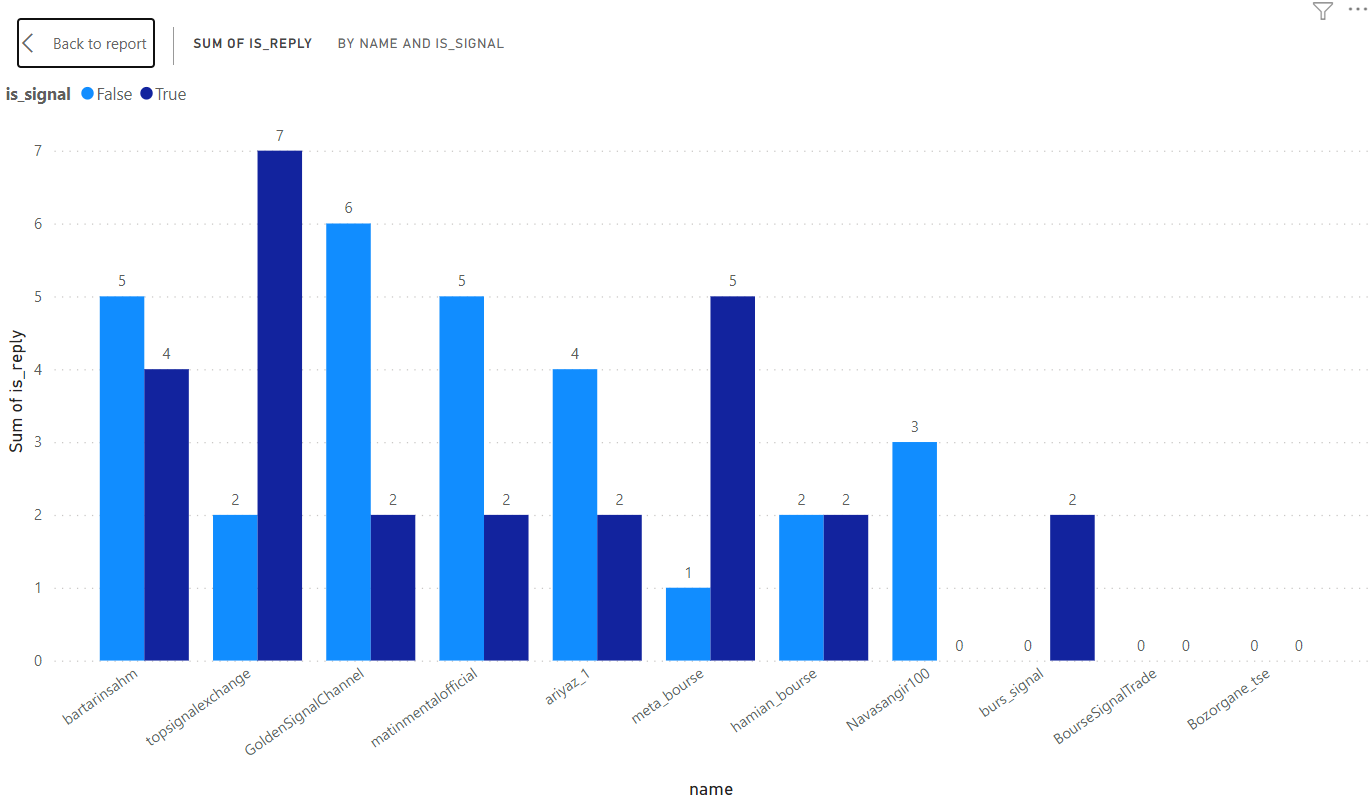
نمودار ۵-۵

## تعداد پاسخ‌های سیگنال‌ دهی شده

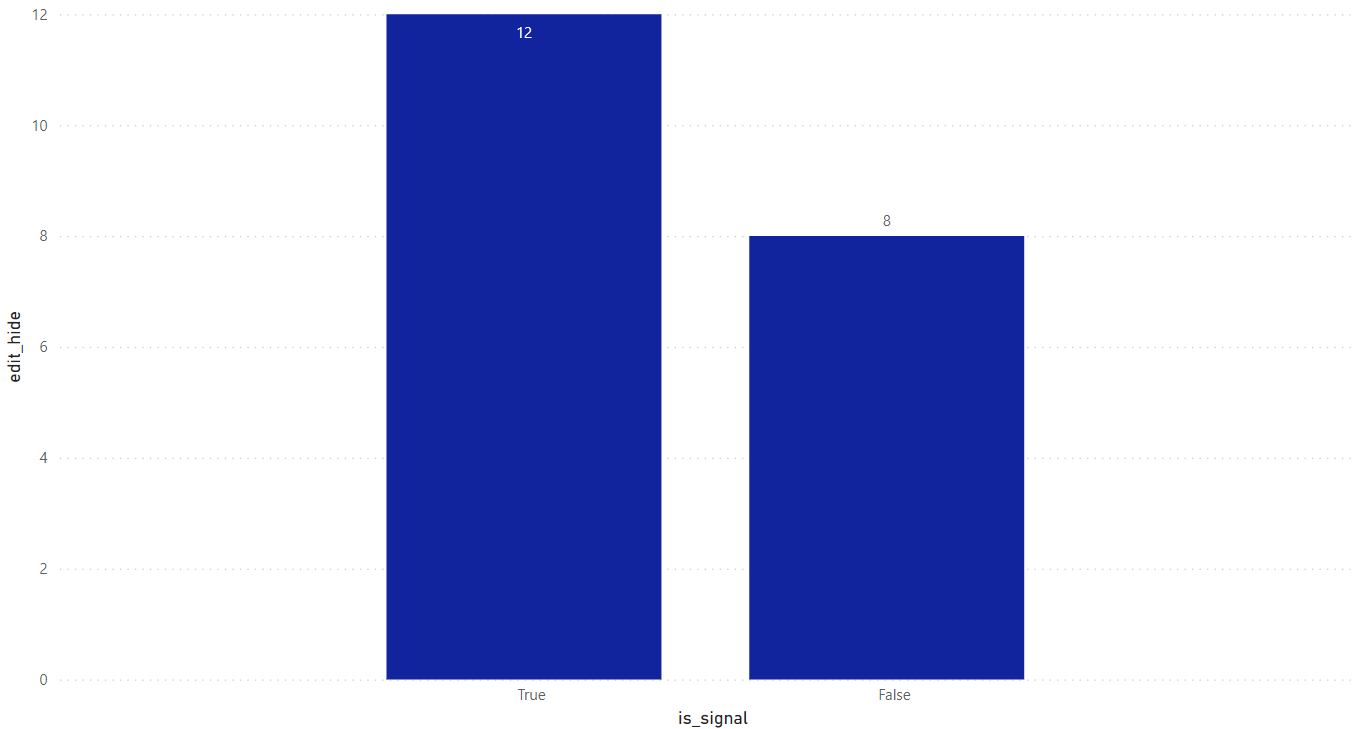
گاهی در کانال‌های تلگرامی عکس، نموداری یا پیامی را ارسال می‌کنند و در پاسخ به محتوای مربوطه سیگنال دهی انجام می‌دهند. درنتیجه بررسی پیام‌هایی که پاسخ به محتوایی هستند و سیگنال دهی انجام می‌دهند اطلاعات مفیدی را در اختیار ما قرار می‌دهد. نمودار ۵-۶ این امر را موردبررسی قرار داده است.

## سیگنال‌ دهی‌هایی که ادیت پیام پنهان‌شده است

ممکن است کانالی سیگنال دهی غلط یا درستی انجام دهد و بعد از گذر زمان پیام مربوطه را تغییر دهد. این در صورتی مشکل ساز است که ادمین کانال در تنظیمات تغییراتی را ایجاد کرده باشد تا کاربران متوجه تغییر داده شدن پیام نشوند. به‌عنوان‌مثال ممکن است سیگنال دهی غلطی انجام دهد و بعد از مشخص شدن نتیجه واقعی این سیگنال دهی را تغییر دهد و به کاربران بگوید که سیگنال دهی او صحیح بوده است. نمودار ۵-۷ این موضوع را موردبررسی قرار می‌دهد.



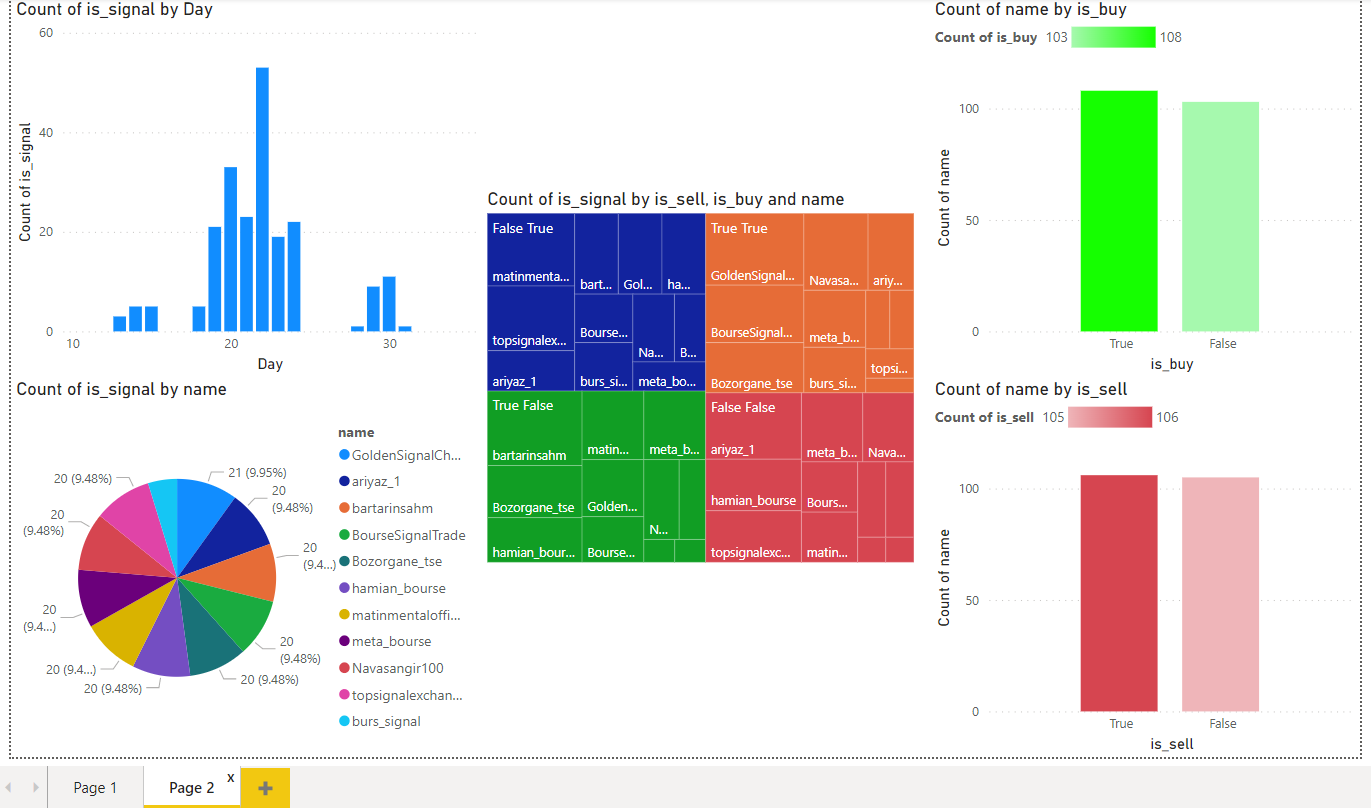
نمودار ۵-۶



نمودار ۵-۷

### نمودار دوم: بررسی دقیق‌تر

نمودار ۵-۸ مواردی را همچون تعداد سیگنال، تعداد سیگنال خرید، تعداد سیگنال فروش، سیگنال‌های فرستاده‌شده در روز‌های خاص و گستردگی سیگنال‌های تحلیل‌شده را به نمایش می‌گذارد.



نمودار ۵-۸

## گستردگی سیگنال دهی در روز‌های مختلف

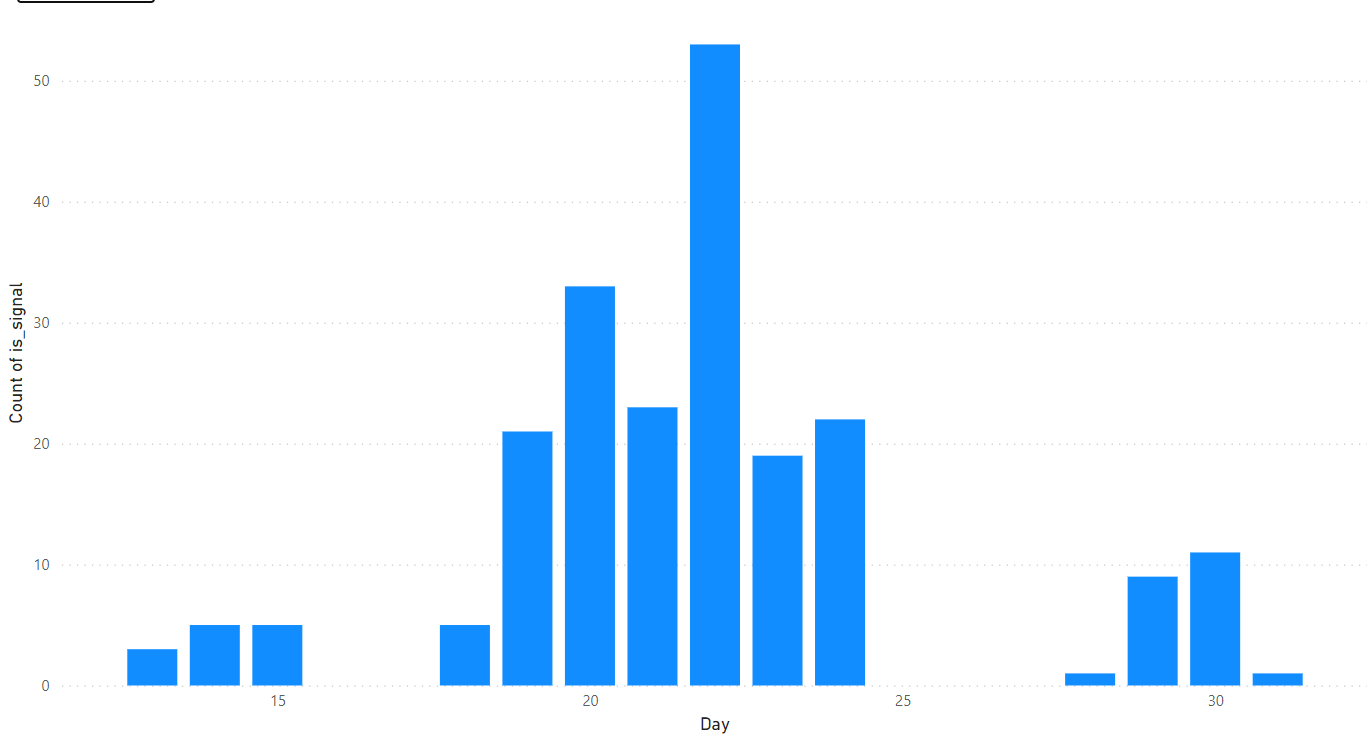
گاهی ممکن است با یک سیگنال دهی تمامی کانال‌ها آن را به اشتراک بگذارند یا به‌صورت بازه‌ای سیگنال دهی زیاد باشد و در آینده کاهش یابد. این امر می‌تواند به علت نشر شدن اطلاعاتی نهانی باشد و باید رصد شود. نمودار ۵-۹ این مقایسه را انجام داده است. همان‌طور که مشاهده می‌کنید در بازه ۲۰ الی ۲۵ام خردادماه سیگنال‌دهی بیشتری صورت گرفته. البته تعطیلات و یا بد بودن شرایط بازار در این موضوع تأثیرگذار می‌باشد.

## سیگنال‌دهی انجام‌شده توسط کانال‌ها

این موضوع که هر کانال چه تعداد سیگنال دهی داشته‌اند در نمودار ۵-۱۰ که نمودار Pie-Chart می‌باشد قابل‌مقایسه است.

## گستردگی سیگنال خرید و سیگنال فروش

همان‌طور که در بخش‌های قبل توضیح دادیم ممکن است بعضی پیام‌ها سخت باشند و تحلیلی دشوار داشته باشند. و یا پیامی بلند باشند که چند سیگنال را در بردارند. گاهی ممکن است کانال متقلبی سیگنالی صحیح با اطلاعات نهانی را در دل سیگنال‌های دیگر جای دهد تا به چشم نیاید و تصادفی تصور شود. بنابراین مهم است تا گستردگی سیگنال خریدوفروش در بین کانال‌ها موردبررسی قرار گیرد تا مواردی ازاین‌دست مشاهده شوند. مقایسه گستردگی در نمودار ۵-۱۱ صورت گرفته است.



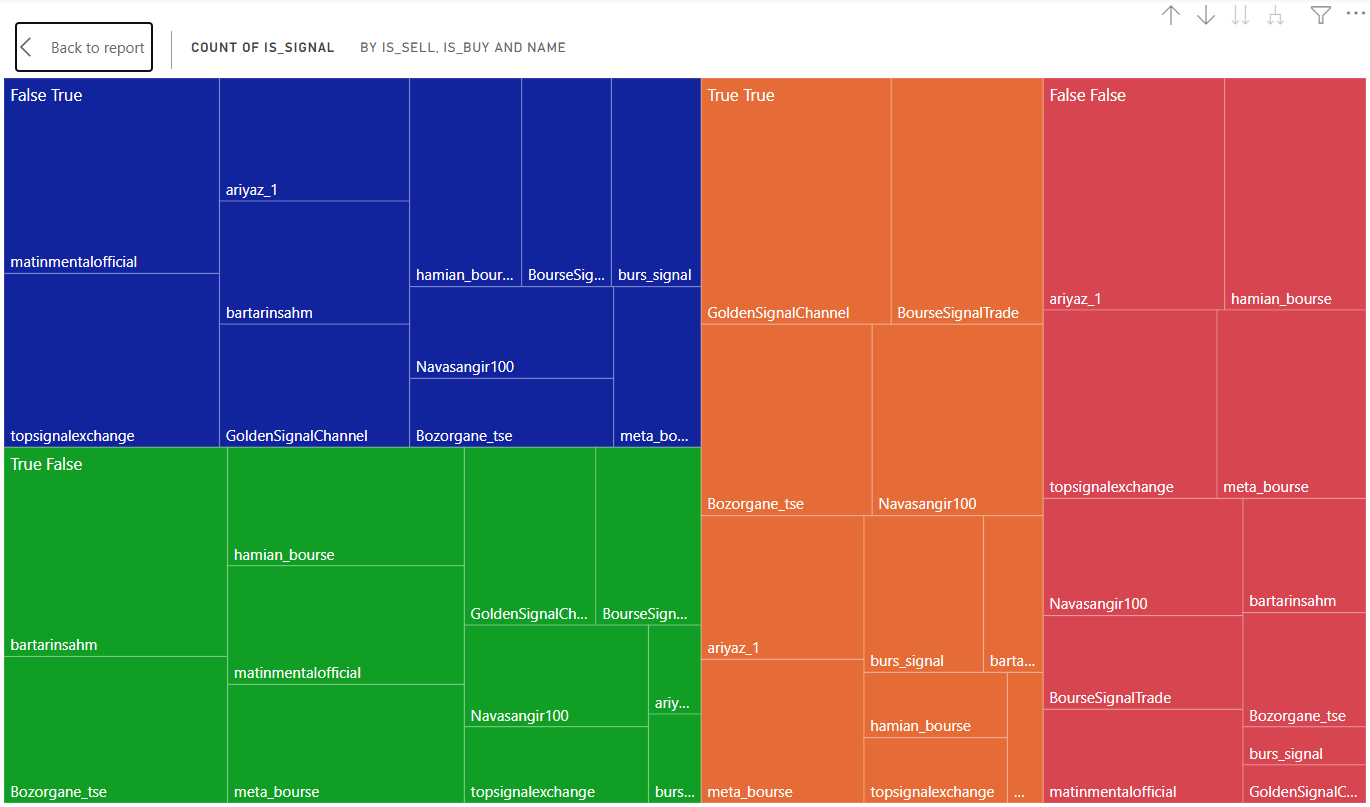
نمودار ۵-۹



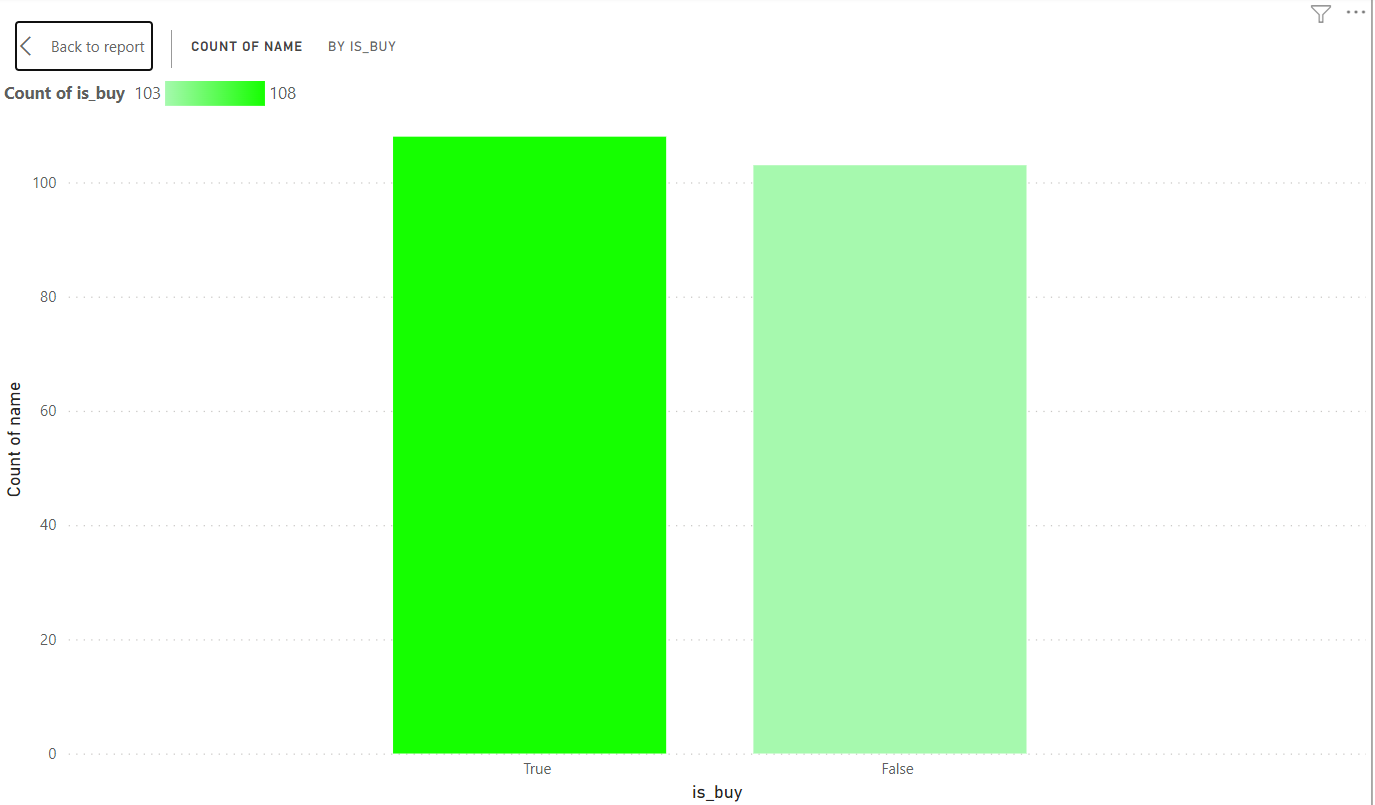
نمودار ۵-۱۰

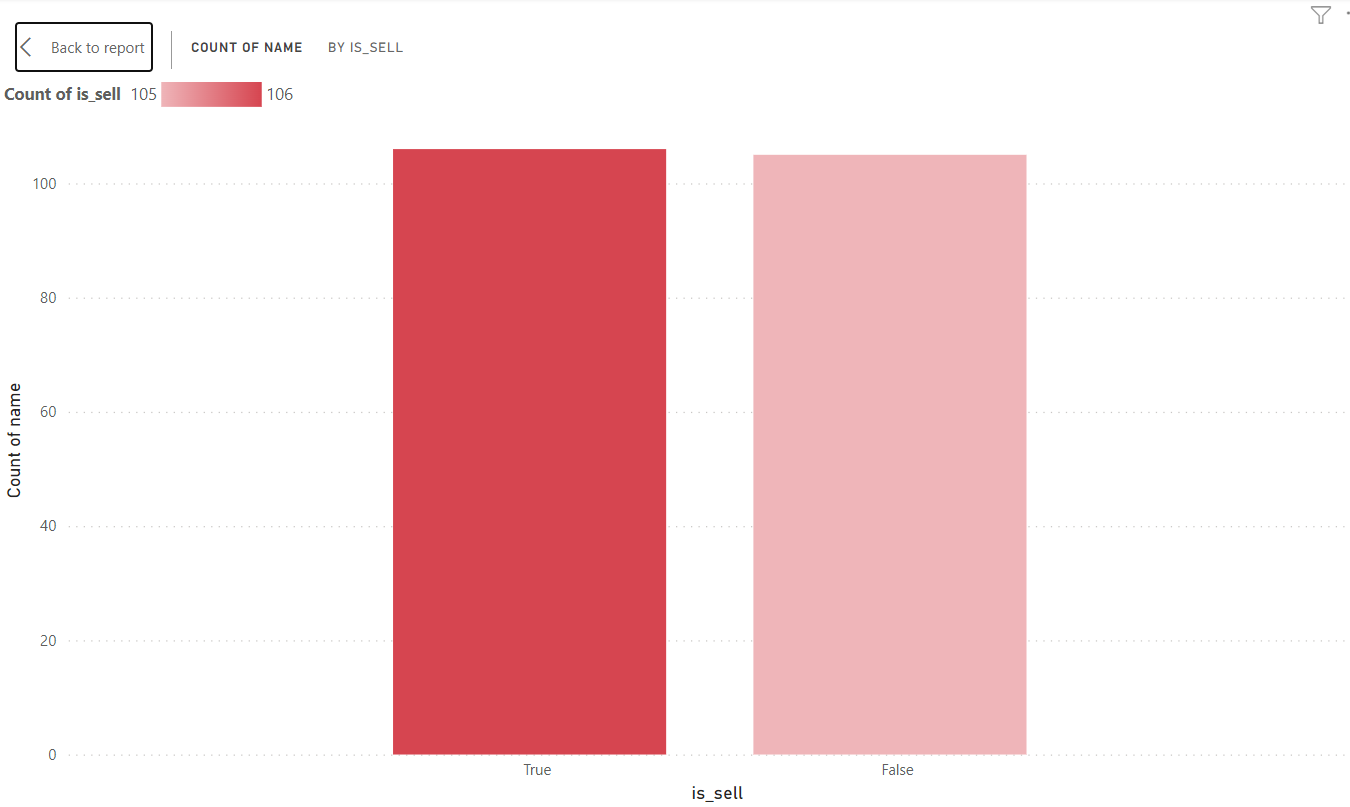
## تعداد سیگنال خرید/فروش کانال

ممکن است با توجه به شرایط بازار بورس اوراق بهادر سیگنال دهی خرید یا فروش بیشتر صورت گیرد. مقایسه تعداد سیگنال‌های خرید در نمودار ۵-۱۲ و مقایسه تعداد سیگنال‌های فروش در نمودار ۵-۱۳ قابل‌مشاهده هستند.



نمودار ۵-۱۱

نمودار ۵-۱۲



نمودار ۵-۱۳

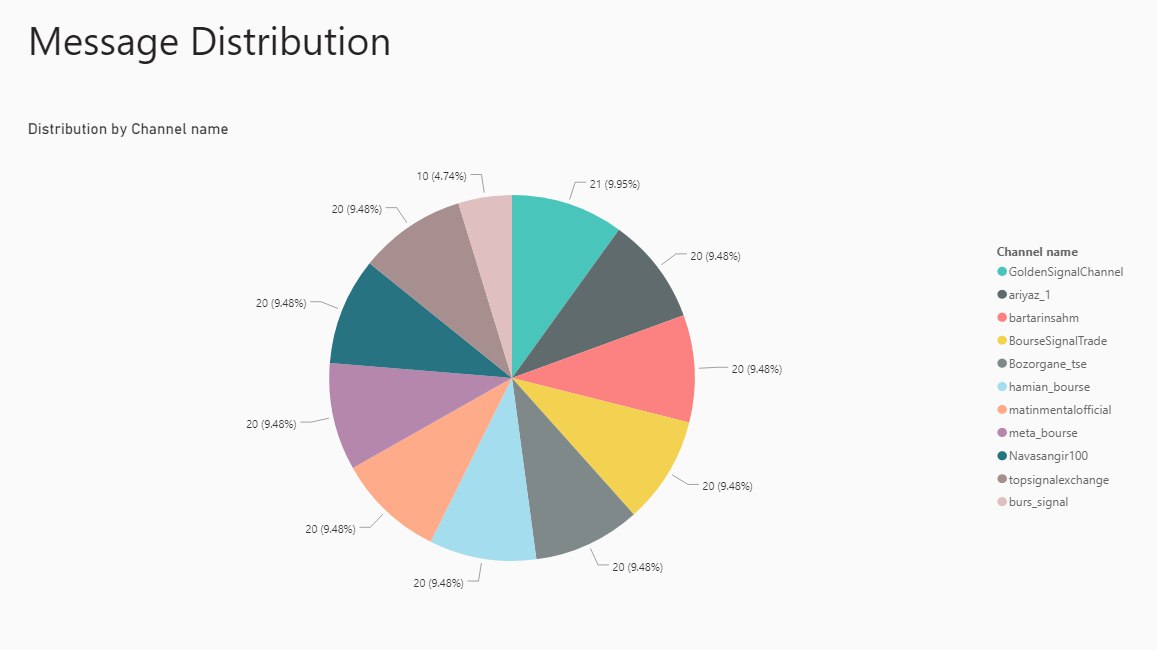
## نمودار‌های تکی

نمودار‌های تکی مواردی‌ را همچون تعداد پیام استخراج‌شده از هر کانال، تعداد سیگنال/غیر سیگنال هر کانال، تعداد سیگنال خریدوفروش و تعداد بازدید رد ساعات مختلف روز را در بردارد. از این نمودار‌ها می‌توان تحلیل‌هایی همچون ساعت اوج فعالیت کانال‌ها و گستردگی سیگنال‌های هر کانال را با یکدیگر مقایسه کرد و اطلاعات مفیدی را درباره کانال بدست آورد. نمودار ۵-۱۴ گستردگی پیام‌های استخراج‌شده را به نمایش می‌گذارد. این نمودار در تمامی نمودار‌های تکی تکرار خواهد شد.

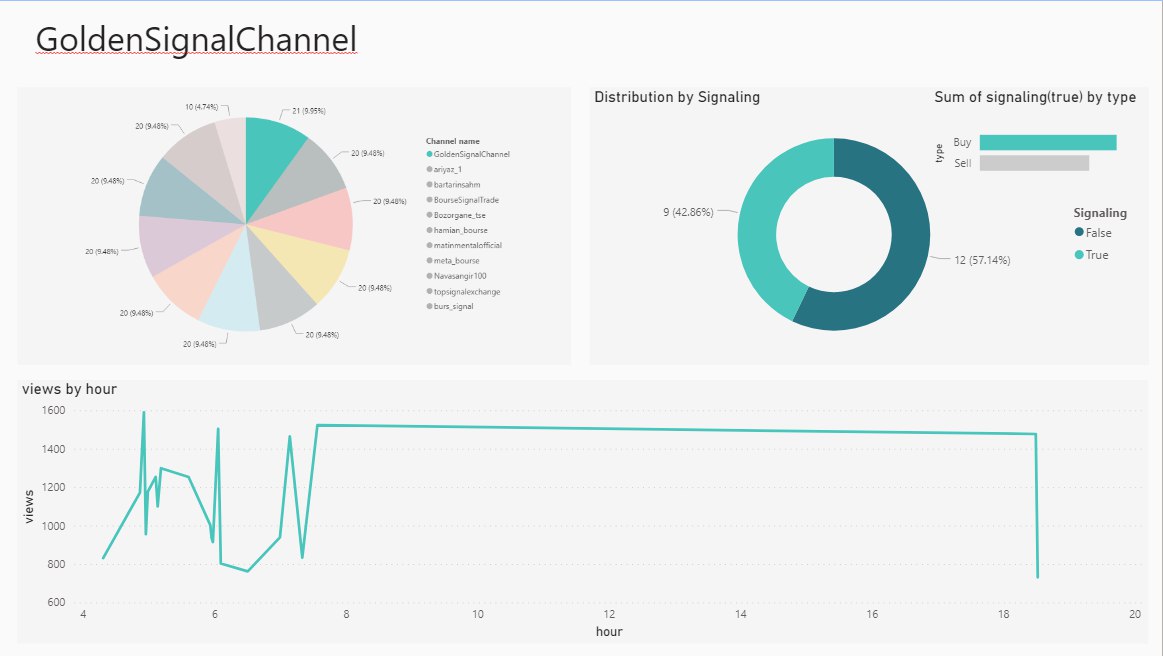
### نمودار تکی مربوط به کانال GoldenSignalChannel

این نمودار (نمودار ۵-۱۵) شامل ۳ نمودار کوچک‌تر می‌باشد. نمودار بالا چپ همان نمودار ۵-۱۴ می‌باشد که بخش مربوط به کانال مربوطه را نشان می‌دهد. نمودار بالا راست تعداد سیگنال‌ها را با تعداد غیر سیگنال‌ها مقایسه می‌کند همچنین یک نمودار افقی وجود دارد که از بین پیام‌های سیگنال تعداد سیگنال‌های خرید را با سیگنال‌های فروش مقایسه می‌کند. نمودار سوم نیز مشخص می‌کند پیام‌های ارسال‌شده در ساعات مختلف روز چقدر بازدید داشته‌اند. این اطلاعات مفیدی همچون ساعت اوج کانال را به نمایش می‌گذارد. همچنین نشان می‌دهد ادمین کانال در چه ساعاتی فعالیت بیشتری دارد.

توضیح تمام نمودار‌های تکی به همین شکل می‌باشد و تنها کانال مربوطه متفاوت است. درنتیجه از زیاده‌گویی پرهیز می‌کنیم و فقط شرح نمودار اول می‌پردازیم.

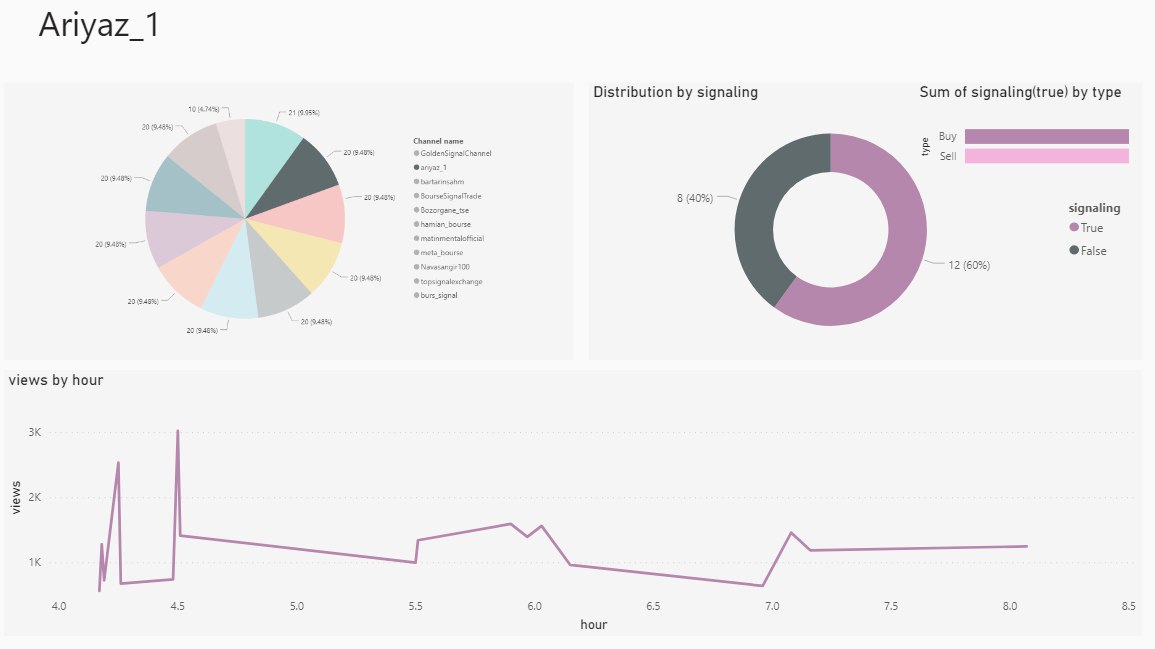


نمودار ۵-۱۴



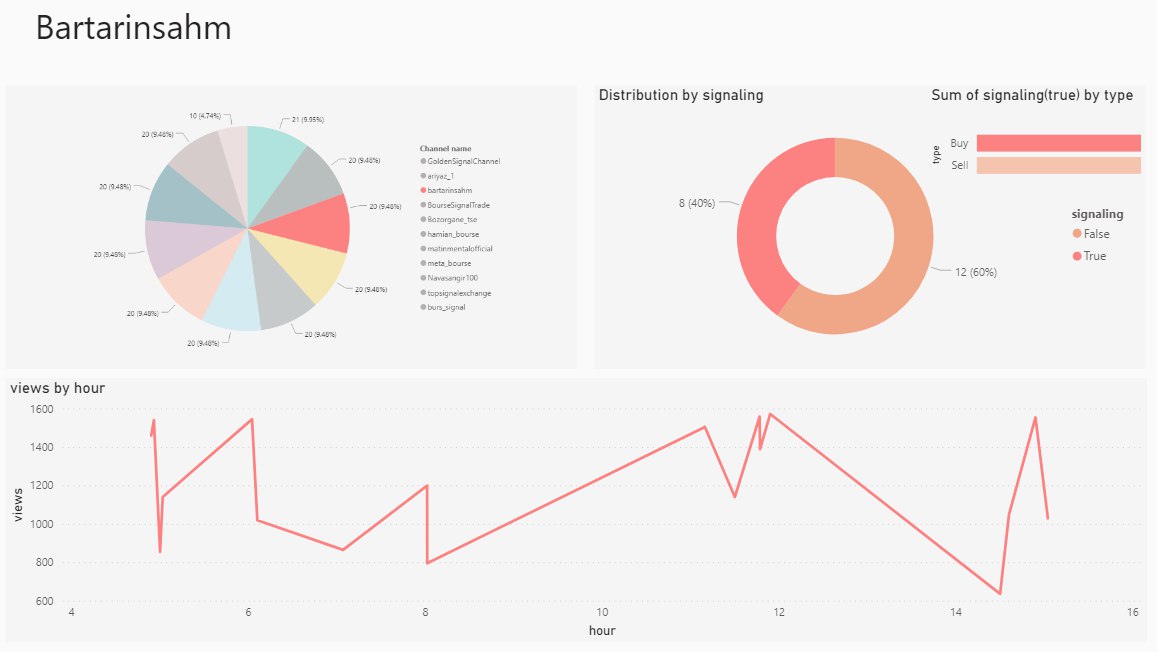
نمودار ۵-۱۵

### نمودار تکی مربوط به کانال Aryaz\_1



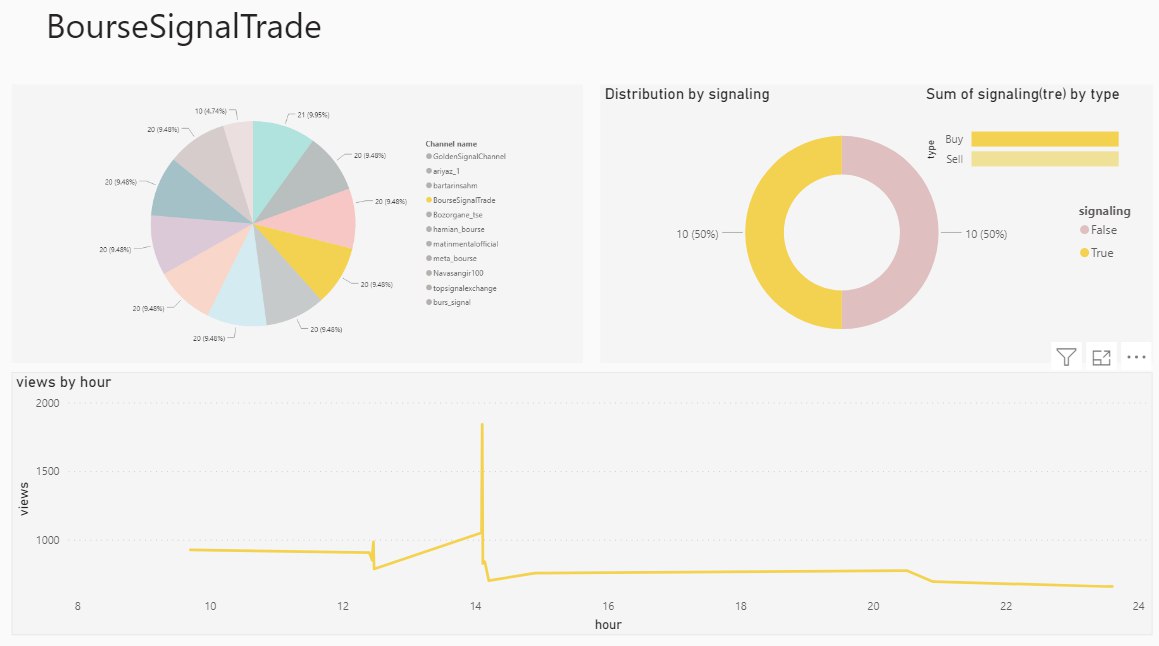
نمودار ۵-۱۶

### نمودار تکی مربوط به کانال Bartarinsahm



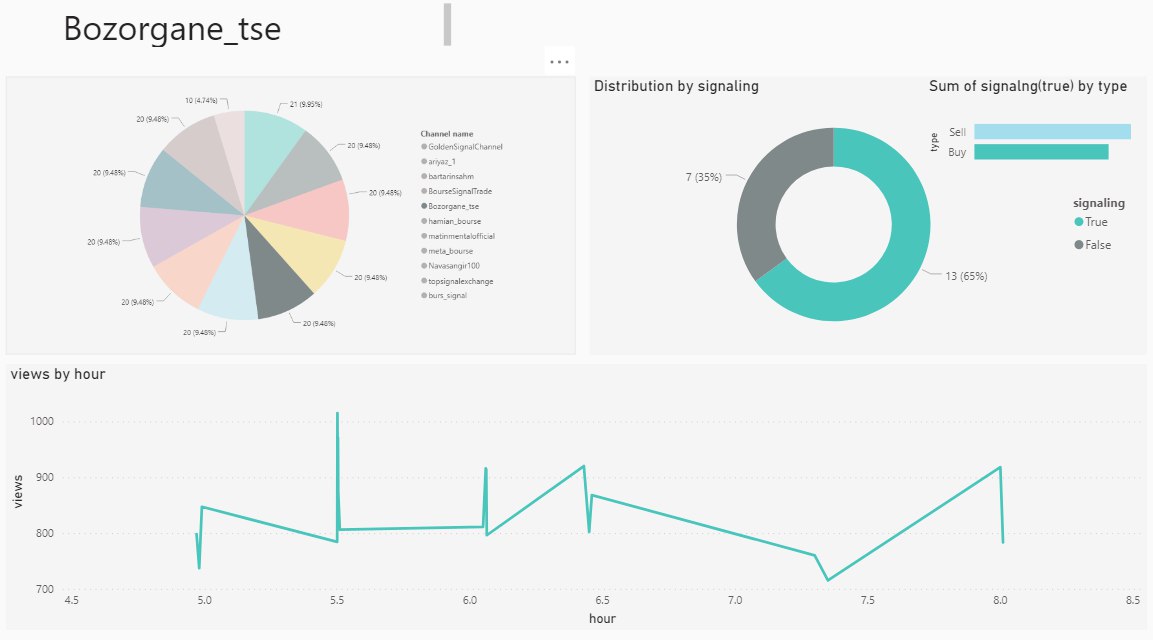
نمودار ۵-۱۷

### نمودار تکی مربوط به کانال BourseSignalTrade



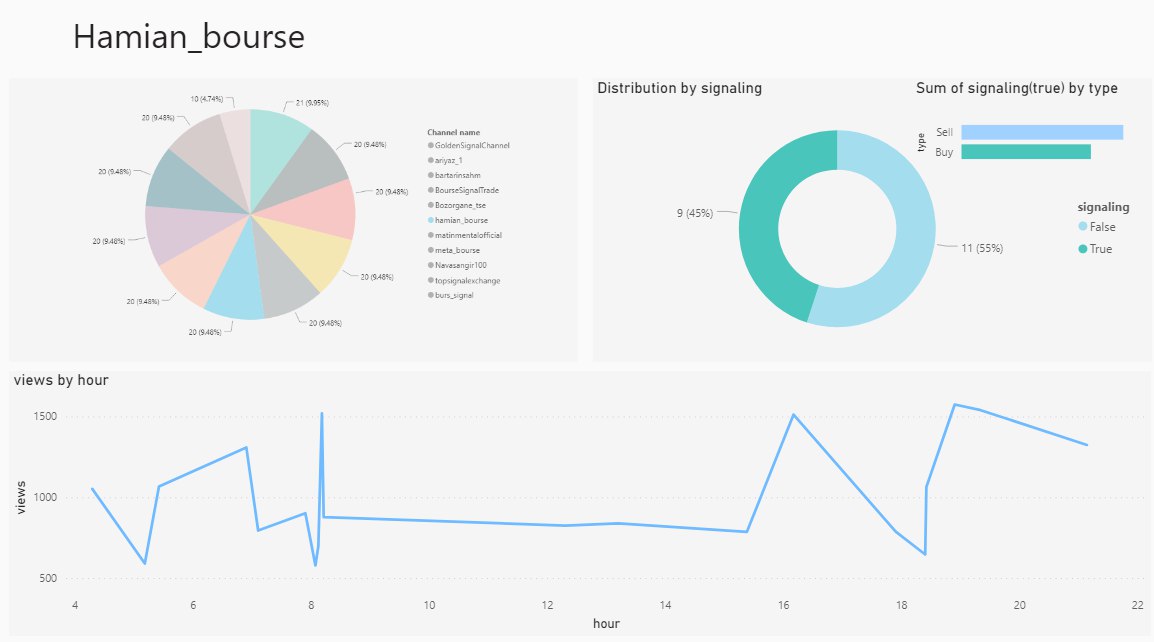
نمودار ۵-۱۸

### نمودار تکی مربوط به کانال Bozorgane\_tse



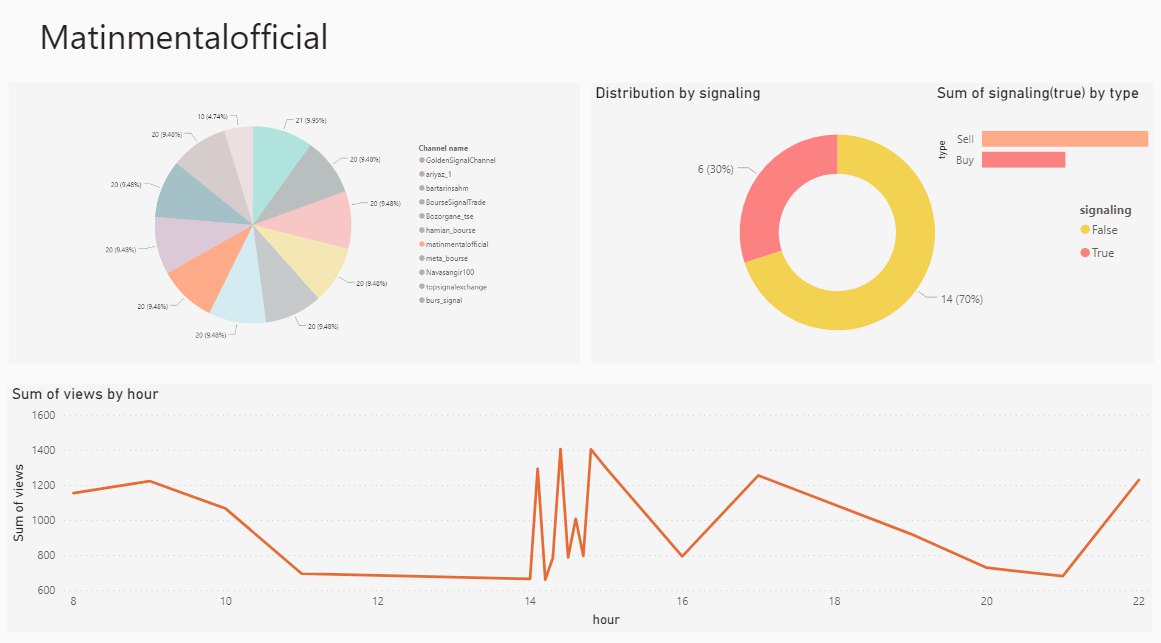
نمودار ۵-۱۹

### نمودار تکی مربوط به کانال Hamian\_bourse



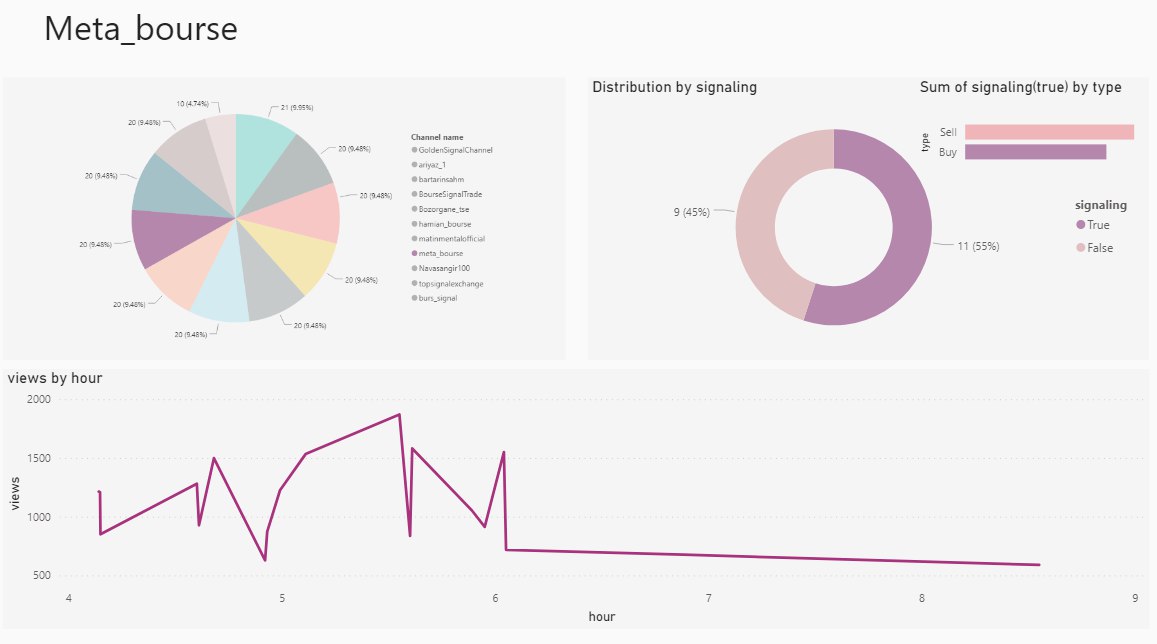
نمودار ۵-۲۰

### نمودار تکی مربوط به کانال Matinmentalofficial



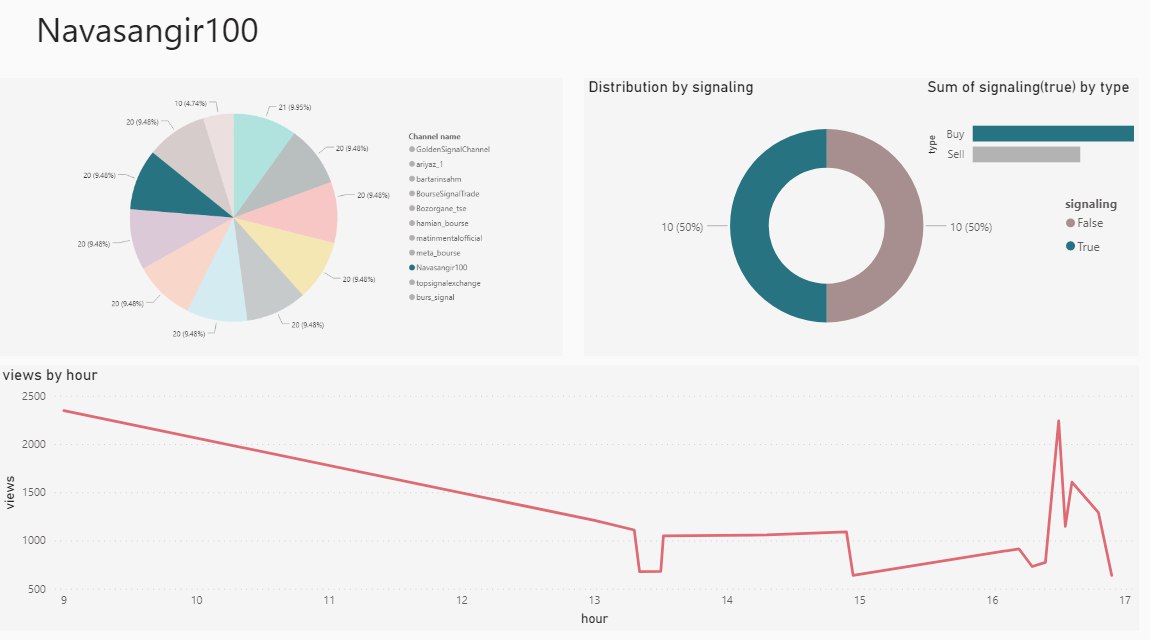
نمودار ۵-۲۱

### نمودار تکی مربوط به کانال Meta\_bourse



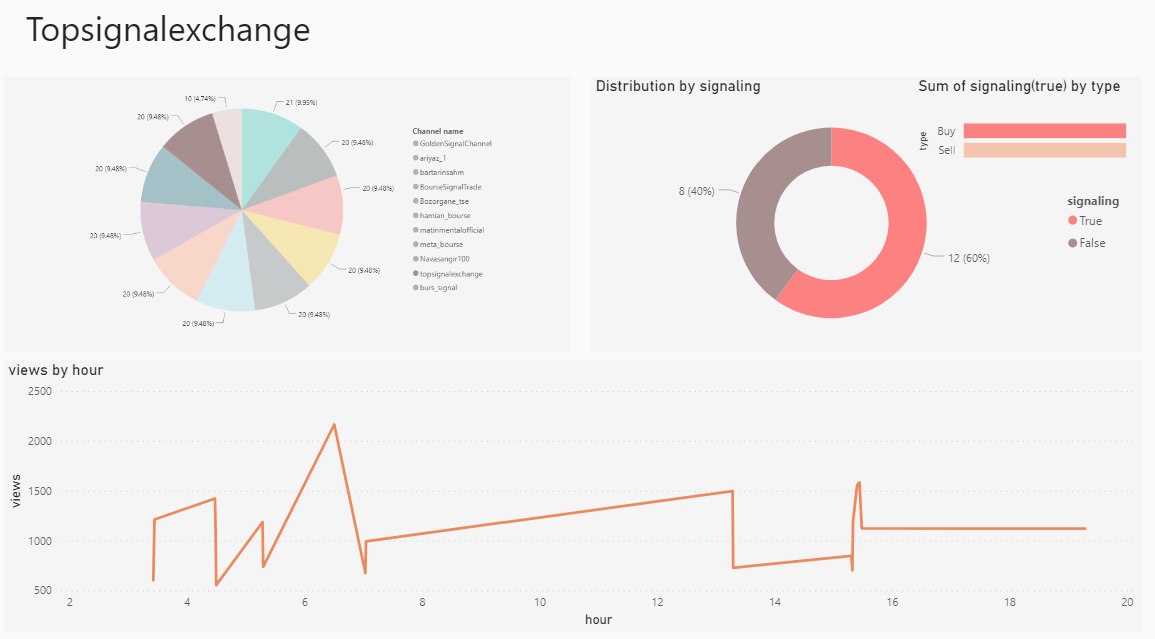
نمودار ۵-۲۲

### نمودار تکی مربوط به کانال Navasangir100

****

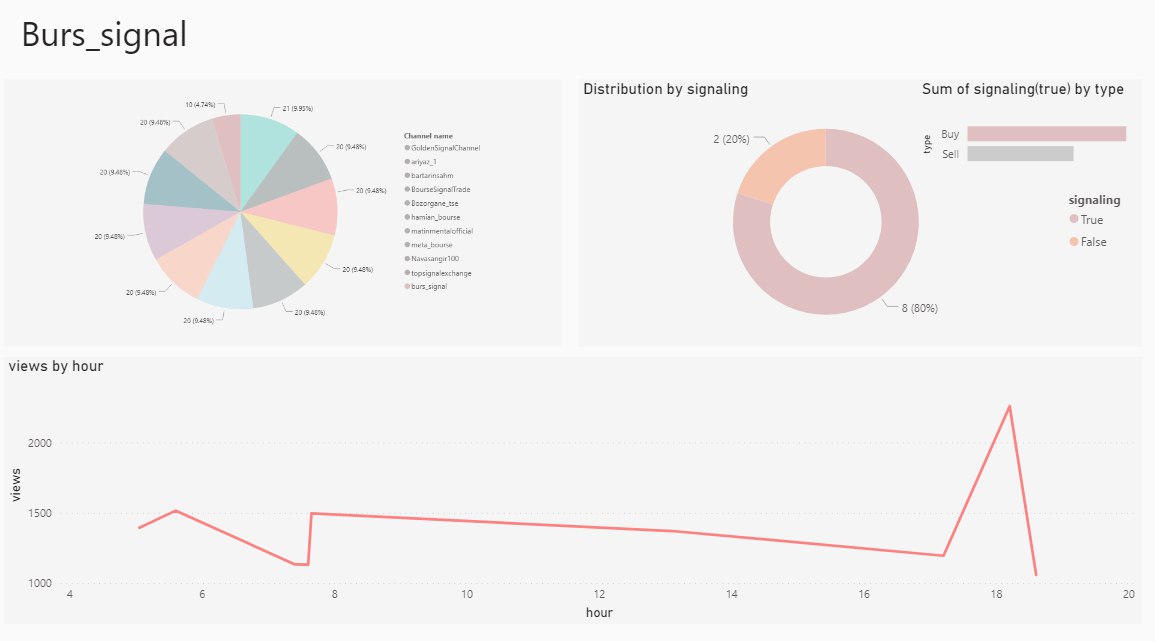
نمودار ۵-۲۳

### نمودار تکی مربوط به کانال TopSignalExchange



نمودار ۵-۲۴

### نمودار تکی مربوط به کانال Burs\_signal



نمودار ۵-۲۵

# فصل ششم جمع‌بندي و نتيجه‌گيري و پیشنهاداتجمع‌بندي و نتيجه‌گيري

در این پروژه اثبات کردیم که امکان پایش رویداد‌های فضای مجازی امکان‌پذیر می‌باشد. همچنین چندین روش من‌جمله تحلیل گروهی و تکی را نیز معرفی کردیم. این پروژه به علت اینکه در آینده دور می‌تواند بیشتر تخلفات را در بورس اوراق بهادار تشخصی دهد اهمیت بسزایی دارد. همچنین با انجام پژوهش‌ها و پرو‌ژه‌های آینده می‌توان این پروژه را گسترش داده و برای تمامی شبکه‌های اجتماعی اعمال کنیم.

**پیشنهادات**

نیازمند به پژوهش‌های بیشتری در این زمینه هستیم چراکه اعتماد عمومی به بورس اوراق بهادار کلید موفقیت این بازار می‌باشد و پژوهش‌هایی ازاین‌دست به محقق شدن اعتماد عمومی کمک بسزایی می‌کند.

**منابع و مراجع**

[1]

“سیگنال‌های لحظه‌ای فارکس، اونس طلا و نفت خام,” *ایران بورس آنلاین*. <https://iranbourseonline.info/forex-signals>

[2]

“Home.” <https://www.stlouisfed.org/>

[3]

“آموزش سرمایه گذاری در بورس,” *خانه سرمایه*. <https://khanesarmaye.com/>

[4]

colorlib, “پرتفوی (به فرانسوی:Portefeuille) یا سبد سهام چیست؟,” *آموزش فارسی*, <http://amoozeshefarsi.com/portfolio/>

[5]

“تابلو خوانی و بازار خوانی بورس چیست - آموزش کامل و حرفه ای تابلو خوانی بورس + PDF -,” *چارت ایران*, 30 می 2020. <https://chartiran.com/2364/tablokhani/>

[6]

Tarnama.ir, “تحلیل ها سیگنال 24,” *بورس 24، آخرین اخبار و تحلیل های بورس تهران | Bourse24*. <https://www.bourse24.ir/articles>

[7]

“چارت ایران - سایت آموزش بورس از مقدماتی تا پیشرفته رایگان,” *چارت ایران*. <https://chartiran.com/>

[8]

“چگونه در بورس سود کنیم؟,” *همشهری آنلاین*, 20 اکتبر 2021. <https://www.hamshahrionline.ir/news/633188/چگونه-در-بورس-سود-کنیم>

[9]

“سامانه سایت کدال چیست و چه کاربردی برای سهامداران دارد؟.” <https://www.sdbroker.ir/article/172/%D8%B3%D8%A7%DB%8C%D8%AA-%DA%A9%D8%AF%D8%A7%D9%84>

[10]

“سایت کدال (codal.ir) چیست؟ آموزش تصویری سایت کدال,” *سیگنال*. <https://isignal.ir/tutorial/معرفی-سایت-کدال/>

[11]

colorlib, “سهام (Share) چیست و چه سهامی در بورس معامله می شود؟,” *آموزش فارسی*, 1 ژوئن 2020. <http://amoozeshefarsi.com/what-is-share/>

[12]

“سهمتو | بهترین های بازار ارز دیجیتال و بورس در یک نگاه,” *سامانه کپی ترید سهمتو*. <https://sahmeto.com>

[13]

“سیگنال بورس چیست ؟ — به زبان ساده (+ فیلم آموزش رایگان),” *فرادرس - مجله‌*, 14 جولای 2020. <https://blog.faradars.org/سیگنال-بورس-چیست/>

[14]

“کارگزاران بورس چه وظایفی برعهده دارند؟,” *آموزش مفید*, 13 آوریل 2020. <https://learning.emofid.com/what_is_brokers_job/>

[15]

“کدال چیست - آموزش کامل سامانه سایت کدال 360 در بورس - آموزش کدال خوانی با فیلم و PDF,” *چارت ایران*, 22 ژوئن 2020. <https://chartiran.com/2817/codal/>

[16]

“یادگیری صفر تا صد آموزش بورس |معرفی نمادها و ارائه تحلیل | سایت سیگنال,” *سیگنال*. <https://isignal.ir/سهام/>

# Abstract

# 

In recent years, the stock market has attracted a lot of attention. The reason for this can be from economic conditions to suitable investment conditions in the stock market. That's why people with less experience in the stock market were looking for people who would be their mentor in this new market.

Many people in social networks with the name of stock market analyst make offers to buy or sell stocks. Sometimes there are people who take advantage of shareholders by abusing "confidential information" and "signaling". This abuse undermines public confidence in the stock market.

In this project, we review several plans to investigate and detect fraud in the stock market. The design proposed in this project is the recent and conventional methods of the digital world. In this project, in addition to the algorithms of these methods, we examine ways to gain people's trust in the stock market.

**Key Words**: Stock market, Signaling, Fraud detection, Confidential information, Monitoring



Amirkabir University of Technology  
(Tehran Polytechnic)

Department of Computer Engineering

Project Report

Monitoring cyber events to detect fraud in stock market

By

Bardia Ardakanian

Mohammadreza Hassanzadeh

Moujan Mirjalili

Farhan Keihan

Supervisor

Dr. Meysam Nazariani

Advisor

Dr. Meysam Nazariani

June & 2022