**Promotional Emails subjects lines analysis**

**Scope of work**

1. I focused on promotional Email subject lines, so I filter out Transactional Emails (emails asking for set password, emails for order tracking, or order confirmation, etc.)
2. I engineered 8 different features ,these features will be used to quantify each email subject ( depending on Length , price ,Tone , personalization , emojii , product,)
3. I constructed a spreadsheet containing 40 different used subject lines (I get these 40 subject lines from 2024 historical dataset )

* Attached a separate word file explaining the Metadata of created features .

1. I choose ( OR % Clean ) and ( CTR % Clean) to be the target key metrics
2. I decide to use the following features from historical data to construct predictive models

* Volume of email sent
* Time ( Day of weak , Month , Quarter )
* Category ( Main / Reminder ) ( Interval or Final )
* Beside 8 engineered features

**---------------------------------------------------------------------------------------------------------------**

**Data exploration insights:**

**Emails subjects line for the highest OR%**

|  |  |
| --- | --- |
| Subject | OR% |
| Canvastavla 100x75cm för bara 199 kr | .157 |
| Dina bilder på Canvastavla eller Fotopresenter från 49,50 kr | .156 |
| Personlig väggdekoration till fantastiska priser! | .148 |
| Premiumtryck upp till 78% rabatt! | .148 |
| WOW! Canvastavla 80x60cm för 189 kr | .145 |
| Hem och livsstil-produkter från 49 kr | .145 |
| Dekorativ inredning för hemmet till de bästa priset | .143 |
| 3 format ➔ 3 dagar ➔ 1 pris | .141 |

**The previous table shows the following characteristics of the highest subject line OR%**

1. It shows price or discount
2. It include the product
3. It did not contains emoji or pictures ( although emoji has a good effect in CTR % )
4. It is short and compact
5. It is not personalized ( although personalization has a good effect on CTR%)
6. It is in formal language

**Email subjects lines for the highest CTR%**

|  |  |
| --- | --- |
| subject | CTR% |
| En hemlig erbjudande bara för dig, skynda! &#x1f381;  Mystery Sale: What's your secret discount? &#x1f92b; | .218 |
| &#x23; Avslöja dolda rabatter: Din mystiska rea väntar på dig!  &#x23; Reveal hidden discounts:Your mystery sale is waiting for you | .189 |
| Otroligt ✨ 3 XXL-format för 249 kr styck  Incredible ✨ 3 XXL sizes for 249 kr each | .184 |
| Sista chansen: Priser från 30,80 kr - Erbjudandet slutar snart! &#x1f3f4;  Last chance: Prices from 30.80 SEK - Offer ends soon! &#x1f3f4; | .169 |
| Använd din kod 39GX32V0 &#x1f525; | Spara nu!  Use your code 39GX32V0 &#x1f525; | Save now! | .149 |

**The previous table shows the following characteristics of the highest subject line CTR%**

1. It is personalized
2. It containing emoji
3. It is in casual language
4. It is in imperative language

**Analyzing the effect of Days of the week.**

Saturdays has the highest CTR % 🡪 .11

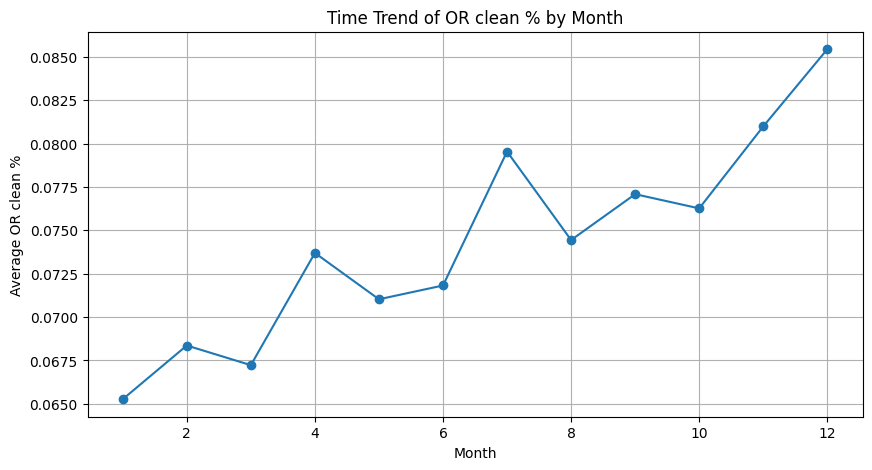
Mondays has the lowest CTR % 🡪 .05

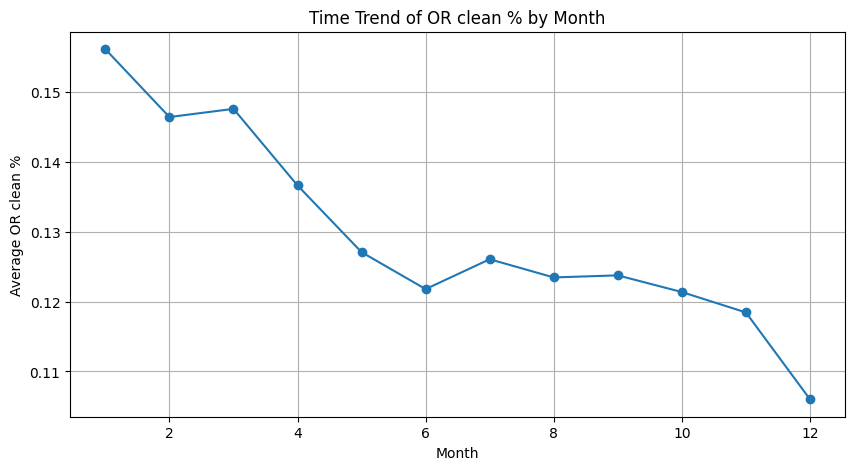
**Analyzing the effect of the month of the year**

**There is a upper trend for the OR% during 2024**

**There is a down side trend for the CTR% during 2024**

**The reasons for these trends needs more investigation**





**Predictive models**

* **I created two simple models to predict both ( OR% and CTR% )**
* **Model accuracy in predicting OR% is R2 = 66% , and there is a good space for improvement if there is more historical data , and more computational power**
* **I believe that having more data about the (the actual content of the emails/data about the recipient of mails) is essential to create more effective CTR% predictive model because predicting CTR% depending on many more factors , but considering the limited features and limited historical data, I created a basic model to predict CTR% , R2 = 51%**
* **Getting more than one year historical data will increase models accuracy**
* **Having access to more computational power will allow me trying more complicated models ( logistic regression / neuron networks )**
* **There is some ambiguity in the provided data, answering some questions about spreadsheet features will increase understanding about different features (difference between OR% and OR clean %?)**
* **There is some other features of subject lines which is not clear ( Color of subject line / Font of subject line / Hour of the Day )**
* **Based of the random forest model , the most important features in predicting subject line OR% is the following**

|  |  |
| --- | --- |
| **Volume of emails sent** | **0.653350** |
| **Month of the year** | **0.114332** |
| **Urgancy of email Tone** | **0.062525** |
| **Imperative Tone in the Email** | **0.030562** |
| **Price or Discount** | **0.024425** |