BrainStation Capstone Sprint 1

Ethan Alwaise

March 8, 2024

The Stack Exchange

A network of Q&A websites for various fields

The Problem

1. Questions:

   - What factors affect the likelihood that a question is answered?

   - What patterns can we observe in posting activity over time?

2. Who should care?

   - Users who want to optimize their questions.

   - Stakeholders in the Stack Exchange.

- What factors affect the likelihood that a question is answered?
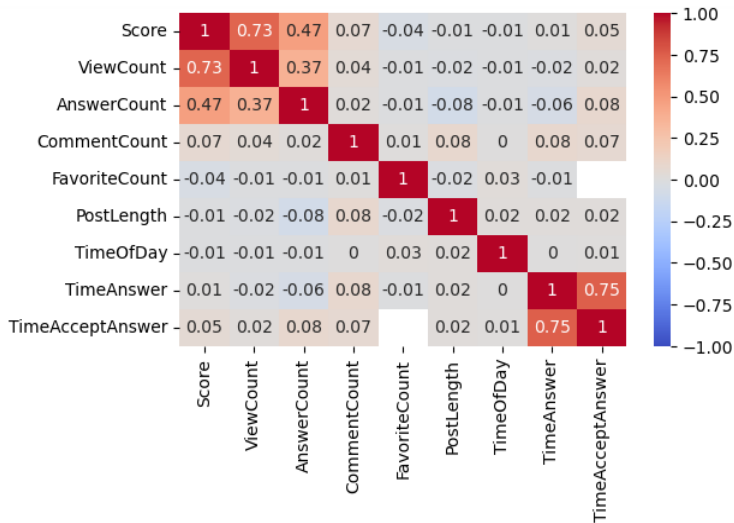
- What patterns can we observe in posting activity over time?

- Binary classification (was the question answered?)

- Regression (how many answers did the question receive?)

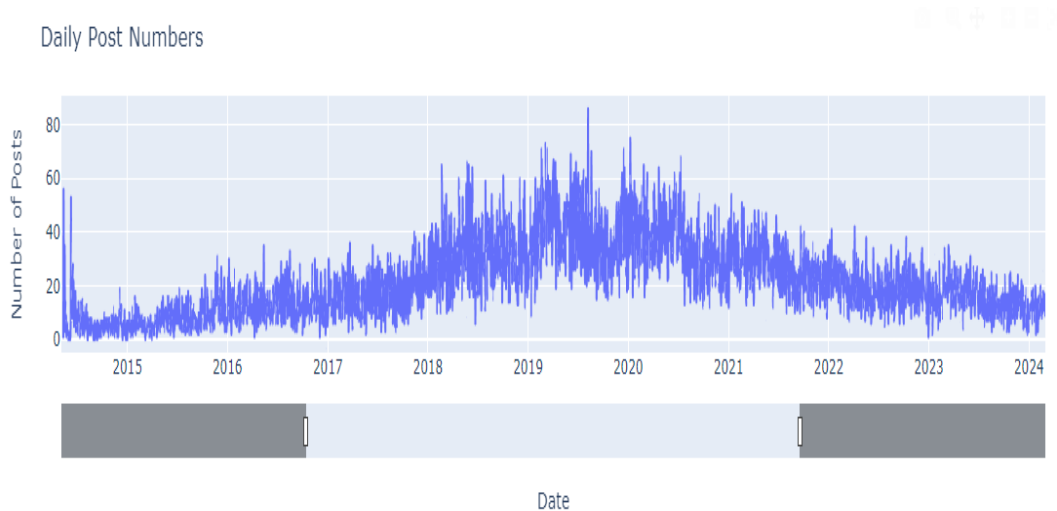- Time series analysis (can we forecast daily post numbers?)

The Data

	PostTypeId	CreationDate	Score	ViewCount	Body	LastActivityDate	Title	Tags	AnswerCount	CommentCount	LastEditDate
0	1	2014-05-13T23:58:30.457	9	959.0	<p>I've always been interested in machine lear...	2014-05-14T00:36:31.077	How can I do simple machine learning without h...	<machine-learning>	1.0	1	None
1	1	2014-05-14T00:11:06.457	4	503.0	<p>As a researcher and instructor, I'm looking...	2014-05-16T13:45:00.237	What open-source books (or other materials) pr...	<education> <open-source>	3.0	4	2014-05-16T13:45:00.237
2	2	2014-05-14T00:36:31.077	5	NaN	<p>Not sure if this fits the scope of this SE,...	2014-05-14T00:36:31.077	None	None	NaN	0	None
3	2	2014-05-14T00:53:43.273	13	NaN	<p>One book that's freely available is "The El...	2014-05-14T00:53:43.273	None	None	NaN	1	None
4	1	2014-05-14T01:25:59.677	26	1925.0	<p>I am sure data science as will be discussed...	2020-08-16T13:01:33.543	Is Data Science the Same as Data Mining?	<data-mining> <definitions>	4.0	1	2014-06-17T16:17:20.473

Correlations

Daily Post Numbers

Date

Next Steps

- Process the text.

- Apply SARIMA to daily post numbers.