

Adverse News Screening for Financial Crime Surveillance

Objective

To develop a system to automatically screen publicly available news articles and identify entities (individuals or corporations) potentially involved in financial crime, scandals, or sanctions.

Below are the steps taken to achieve this:

1. Extracted news articles from CNBC website
2. Identified individuals and corporations mentioned in the news article, tagged them and assigned Wikipedia link to the individual/corporation
3. Used sentiment classification model to classify if news sentiment is positive, neutral or negative
4. Used zero shot classification to label the financial crime topic of the news article
5. Load all output to Tableau dashboard

Benefits

With only adverse financial news, all tagged with their topic, relevance score and a short summary, all stored in a single dashboard for users' quick and easy perusal

- No need to manually scan news website for the relevant articles
- Short summary of article allows for faster reading, URL links still available to read full article
- Relevance score helps users quickly identify which are the more important articles to focus on
- Update watchlist of people or organizations mentioned
- Learn from others' mistakes and review policies and workflows to reduce risks of similar situation occurring

Data Source

- Consumer News and Business Channel (CNBC), a financial and business website, was chosen as a data source as it is mostly unbiased (no political leaning) with numerous free articles.
- 3324 news articles published between Oct 2024 to Jan 2025 were extracted from CNBC for this project.



Identifying Entities

- **SpaCy**, a library for Natural Language Processing, was used to identify Entities in the news article text as:
 - it is fast and easy to use
 - and can be used together with SpaCy Entity Linker
- SpaCy Entity Linker (SpaCy module) matches potential candidates from each sentence (subject, object, prepositional phrase, compounds, etc.) to aliases from Wikidata



SpaCy Entity Linker output entity is not standardised. Hence for each distinct URL, use the entity form that occurs the most for that URL. Use 'Mark Zuckerberg' for this example.

A	B	C	D
entity	entity	description	wikipedia_url
Mark Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Mark Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Mark Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Mark Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Mark Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Zuckerberg	ORG	American internet entrepren	https://www.wikidata.org/wiki/Q36215
Zuckerberg	PERSON	American internet entrepren	https://www.wikidata.org/wiki/Q36215

Adverse News Classification

ML model was used to classify whether news article contains adverse news related to financial crimes, scandals or sanctions

Possible Approach

- Manually label news article by their topic and train ML model on them
 - **Labour intensive**
- Use pre-trained ML models to do the classification, possible ways:
 - A. Adverse news can be considered as negative sentiment, we can use **sentiment classification model to classify adverse news**
 - B. **Defining possible financial crime categories** and inputting them into **zero shot classification models**
 - C. Can **also input categories ['positive', 'neutral', 'negative'] to zero shot classification model** to do sentiment classification

Adverse News Classification

Use both method A and B to identify adverse financial crimes as it is possible article could mention something crime related but the sentiment is positive:

*U.S. blockchain startup Ripple launches new services aimed at helping banks and fintech firms store digital assets. New features include pre-configured operational and policy settings, monitoring of **anti-money laundering risks**, and a new user interface that's easier to use and engage etc.*

This would be labeled anti-money laundering by zero shot classification model but positive sentiment by sentiment classification model

Adverse News Classification

Models tested

- Pre-trained models from Hugging Face were used for classification
- They were selected based on their usage (most likes/downloads) as well as whether they were trained on financial related datasets.



Sentiment Classification	Zero Shot Classification
ProsusAI/finbert	facebook/bart-large-mnli
yiyanhkhust/finbert-tone	MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli
mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis	

Adverse News Classification

Model Selection and Validation

- Create a small labeled dataset and fit into, for example, all possible sentiment classification models
- Evaluate models based on their precision, recall and f1 score
- Use best sentiment model to fit CNBC news dataset
- Randomly sample some articles from CNBC news dataset and check the predicted labels against the article (check if the predicted label make sense)

Adverse News Classification

Topic Modeling Results

- For financial crime topic modeling, dataset with 75 news article extracted was created and labeled as 'fraud', 'tax evasion' and 'non financial crimes'.
- Categories to classify the articles to are: 'fraud', 'scam', 'tax evasion', 'other financial crimes' and 'non financial crimes'.
- The predicted topic that has the highest score is chosen to represent the topic of that article

Adverse News Classification

facebook/bart-large-mnli

	precision	recall	f1-score	support
fraud	0.79	0.79	0.79	28
non financial crime	1.00	0.52	0.68	33
other financial crime	0.00	0.00	0.00	0
scam	0.00	0.00	0.00	0
tax evasion	0.89	0.57	0.70	14
accuracy			0.63	75
macro avg	0.53	0.37	0.43	75
weighted avg	0.90	0.63	0.72	75

MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli

	precision	recall	f1-score	support
fraud	0.79	0.68	0.73	28
non financial crime	0.96	0.76	0.85	33
other financial crime	0.00	0.00	0.00	0
scam	0.00	0.00	0.00	0
tax evasion	1.00	0.43	0.60	14
accuracy			0.67	75
macro avg	0.55	0.37	0.44	75
weighted avg	0.91	0.67	0.76	75

MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli generally performed better (besides the recall for tax evasion)

Adverse News Classification

Sentiment Classification Model Results

Same dataset from the topic modeling is used but now articles are classified as negative, neutral or positive.

ProsusAI/finbert

	precision	recall	f1-score	support
negative	0.94	0.76	0.84	42
neutral	0.52	0.68	0.59	19
positive	0.62	0.71	0.67	14
accuracy			0.73	75
macro avg	0.70	0.72	0.70	75
weighted avg	0.78	0.73	0.75	75

mrm8488/distilroberta-finetuned-financial-news-sentiment-analysis

	precision	recall	f1-score	support
negative	0.80	0.19	0.31	42
neutral	0.16	0.32	0.21	19
positive	0.44	0.86	0.59	14
accuracy			0.35	75
macro avg	0.47	0.45	0.37	75
weighted avg	0.57	0.35	0.33	75

yyianghkust/finbert-tone

	precision	recall	f1-score	support
negative	0.95	0.45	0.61	42
neutral	0.18	0.26	0.21	19
positive	0.41	0.79	0.54	14
accuracy			0.47	75
macro avg	0.51	0.50	0.45	75
weighted avg	0.65	0.47	0.50	75

ProsusAI/finbert outperformed the rest of the models in all metrics, with overall accuracy at 73%.

Adverse News Classification

Final modeling Steps

- Load CNBC news dataset
- Chunk articles that are more than 1024 tokens
- Fit article/chunk to summarisation model **sshleifer/distilbart-cnn-12-6** from Hugging Face to get text with max 514 tokens
- Fit summarised text to **MoritzLaurer/DeBERTa-v3-base-mnli-fever-anli**
- Fit summarised text to **ProsusAI/finbert**
- Sample some of the CNBC news articles to check models' performance

Relevance Scoring

To give a score to the topic predicted by model as to how relevant the news is to financial crime.

Possible Calculation

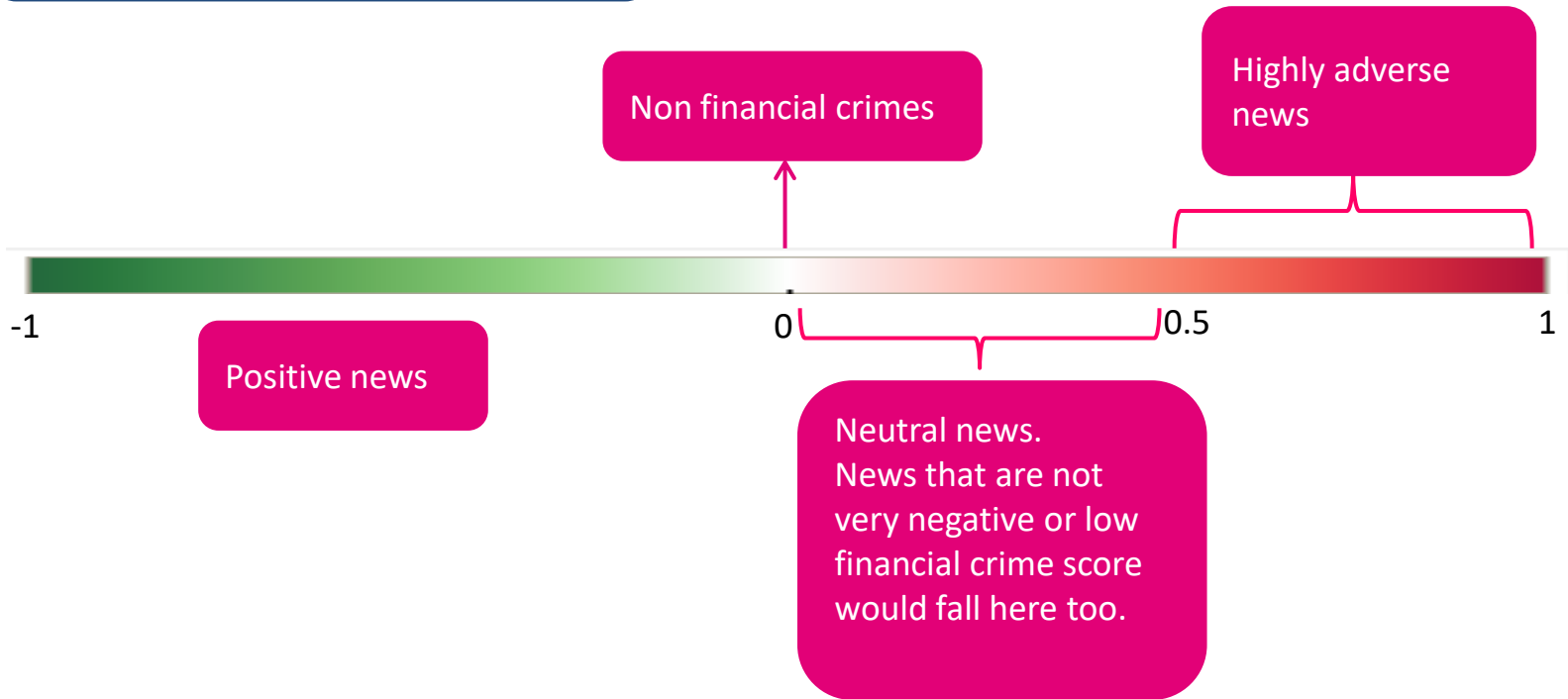
$$\text{Reputation of News website} * \text{financial crime score from model} * \text{sentiment classification score from model}$$

Reputation of news website: Take values 1 to 10 inclusive, where 10 is most reputable and 1 is least reputable.

Financial crime score from model: Takes values 0 to 1, where 1 is most likely to belong to that crime. If the topic belongs to non financial crime, set this score to 0.

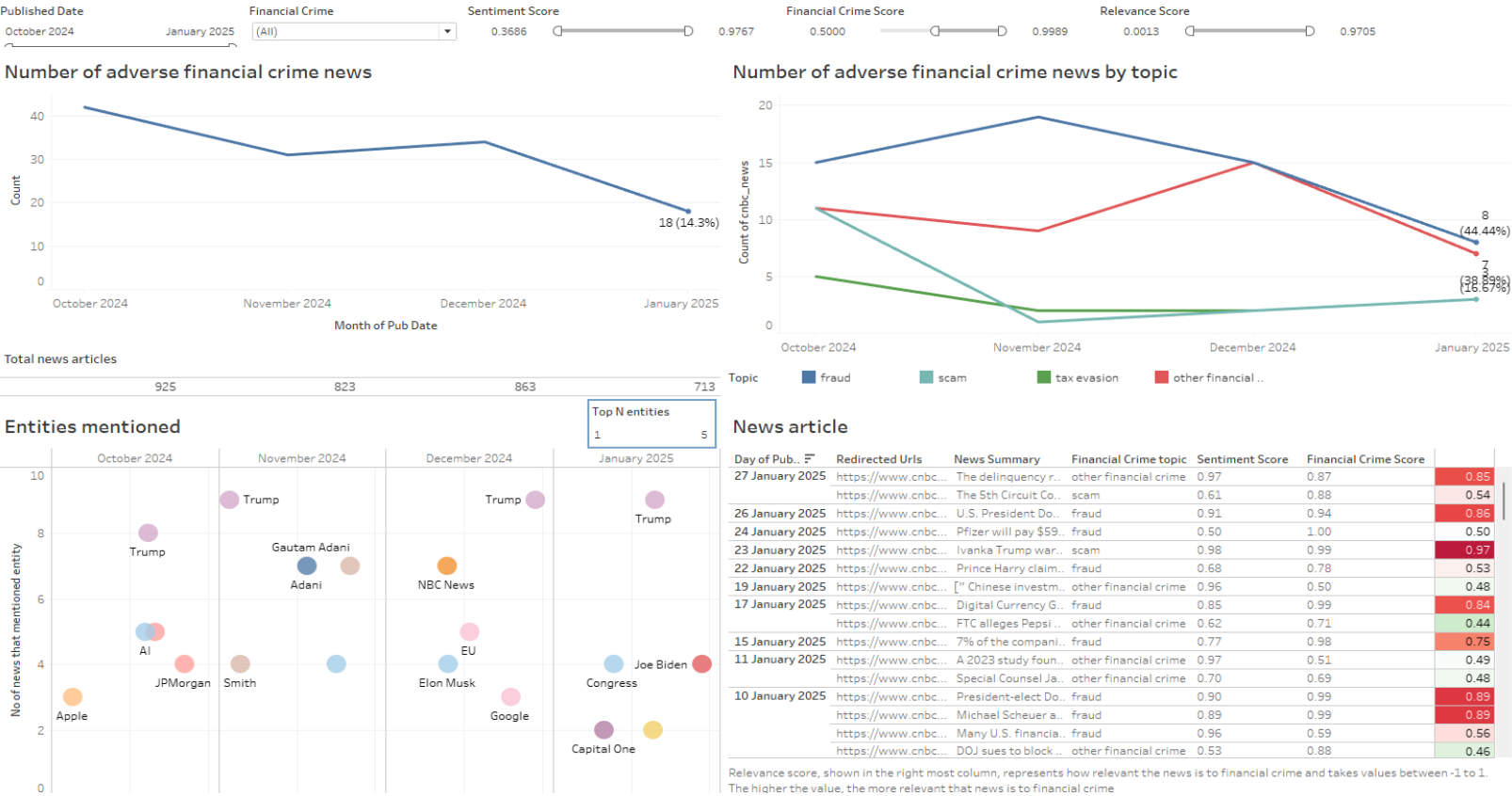
Sentiment classification score: Takes values -1 to 1, where news that are classified as positive shall take the negative of score from sentiment classification model, news that are neutral shall take 0.5 * score from sentiment classification model and news that are negative shall take actual score from sentiment classification model.

Relevance Scoring



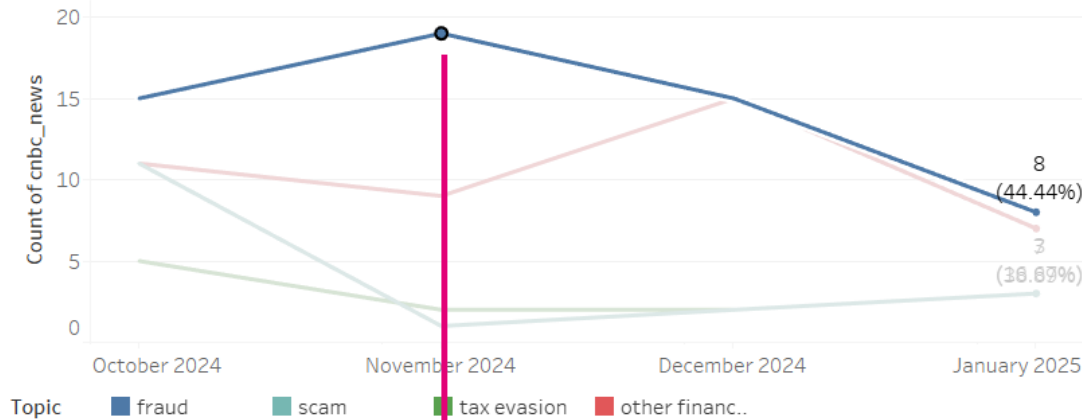
Final output is a dashboard where user can identify trends as well as drill down into points of interest in the data

Financial Crime Dashboard



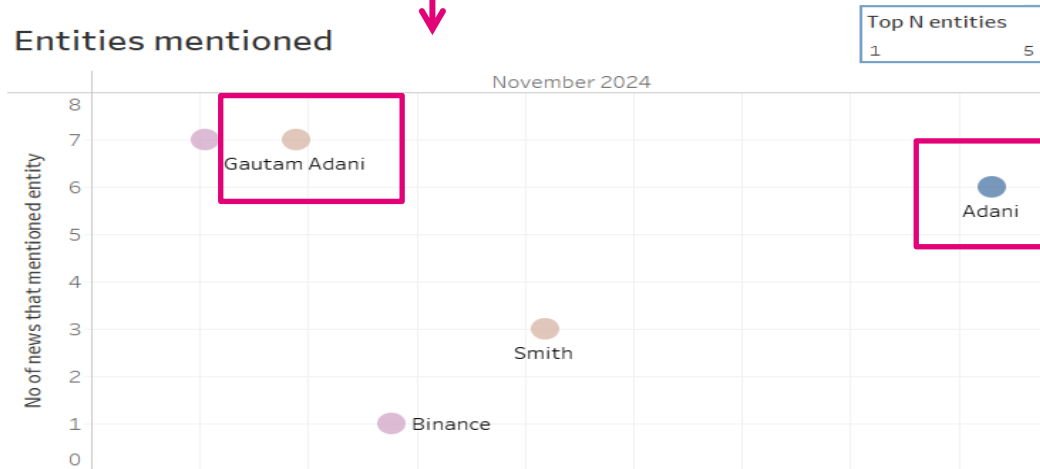
Adani Fraud Charge

Number of adverse financial crime news by topic



High number of fraud cases in Nov 2024, click on line to filter other charts for fraud cases

Entities mentioned

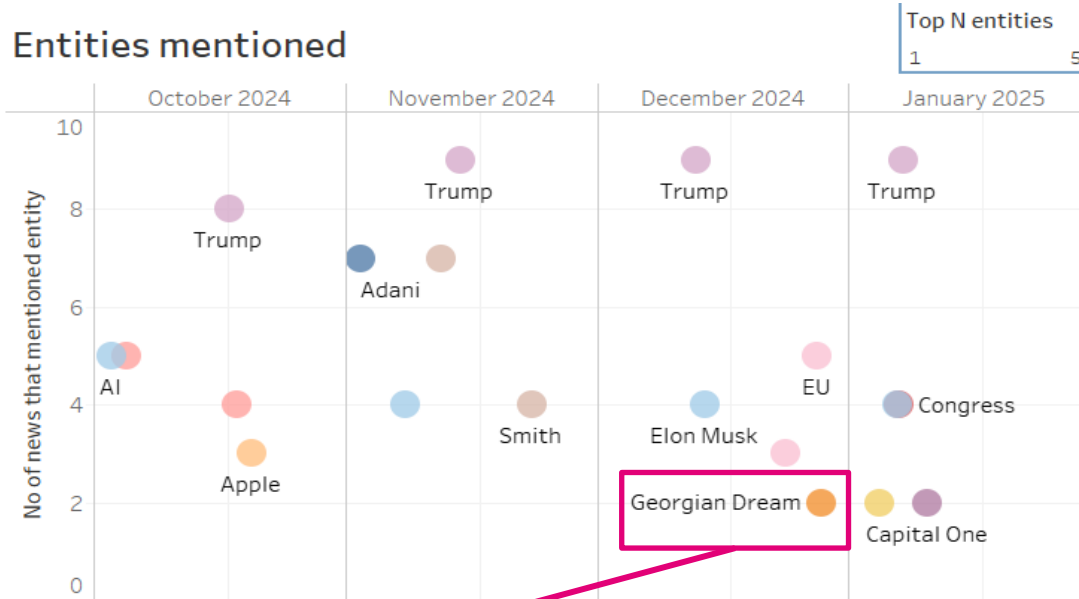


Adani Group and their Group Chair Gautam Adani were allegedly involved in an extensive bribery and fraud scheme, with the bribes paid to obtain solar energy contracts

DBS might need to closely monitor transaction activities of persons tied to Adani Group as well as loans and transactions of solar energy companies in India
DBS can freeze suspicious accounts and transactions for investigations

Sanction on Georgia's ex-prime minister

Entities mentioned



✓ Keep Only ○ Exclude 🔍 📄

President Salome Zourabichvili says Mikheil Kavelashvili has no legitimacy as president . She says he was not duly picked, as the lawmakers who chose him were elected in an October parliamentary election marked by fraud . Opposition parties say that the October election was free and fair . The U.S. imposed sanctions on Bidzina Ivanishvili, a reclusive billionaire ex-prime minister who is widely seen as Georgia's de facto leader . The confrontation comes amid a month of protests sparked by the suspension of EU accession talks .

Exploring the Entities chart, an interesting one to further investigate is 'Georgian Dream'

Filtering news based on 'Georgian Dream', we see Georgian Dream party won parliamentary election marked by Fraud, and its founder Bidzina Ivanishvili has been sanctioned by US.

DBS might need to freeze Bidzina Ivanishvili account
DBS can closely monitor and investigate transactions made by entities with links to Bidzina Ivanishvili

Fake Crypto coin



Filter for relevance score more than 0.9 to get news articles that are more relevant to financial crimes

Article on fake crypto coin called \$IVANKA was being promoted and posed risks to investors

News article

Day of P..	Redirected Urls	News Summary	Financial Crime ..	Sentiment Sc..	Financial Crime S..	
23 January ..	https://www.c..	Ivanka Trump ..	scam	0.98	0.99	0.97
29 Decembe..	https://www.c..	President Sale.. fraud		0.97	1.00	0.97
6 December ..	https://www.c..	Ivanka Trump warns against a fake crypto coin called 'Ivanka Trump' or '\$IVANKA' She says the fake coin is being promoted without her consent or approval . Her warnings come after meme tokens were launched over the weekend called \$Trump and \$MELANIA . The \$Trump coin is now trading at about \$37, a drop of 50% from its peak shortly after launch . The first lady's first-lady token is trading at \$2.84, down 80% from her weekend high .				
1 December ..	https://www.c..					
22 Novembe..	https://www.c..					
21 November 2024	https://www.c..					
20 Novembe..	https://www.c..	Adani Group c..	fraud	0.97	0.99	0.96
15 Novembe..	https://www.c..	Financial firm ..	fraud	0.96	0.95	0.91
11 Novembe..	https://www.c..	FTX filed a law..	fraud	0.95	0.99	0.95
28 October	https://www.c..	Stefan Kaufm..	fraud	0.96	1.00	0.95
2024	https://www.c..	Kristo Käärma..	tax evasion	0.94	1.00	0.94
27 October ..	https://www.c..	Prime Ministe..	other financial c..	0.95	0.96	0.91
11 October ..	https://www.c..	Prime Ministe..	other financial c..	0.97	0.97	0.95

Relevance score, shown in the right most column, represents how relevant the news is to financial crime and takes values between -1 to 1. The higher the value, the more relevant that news is to financial crime

DBS could consider putting up a warning to customers who are transferring monies to crypto exchanges to be wary of fake crypto coins circulating to increase awareness.

DBS could also brief the relationship managers to be aware and inform clients the risks of investing in \$IVANKA.

Q&A