

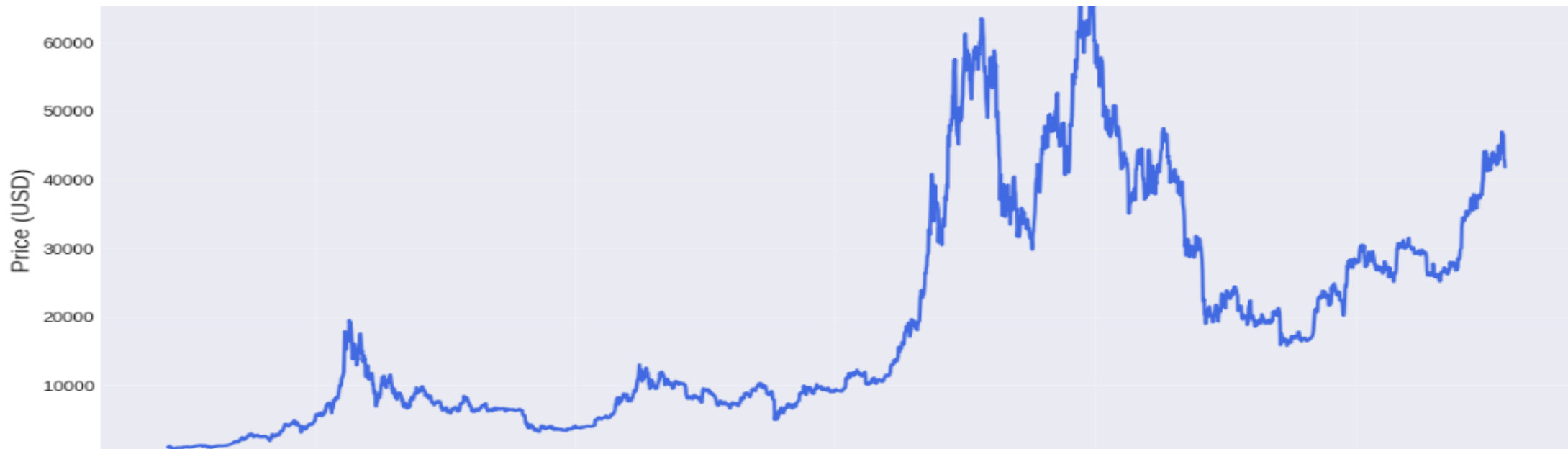
# Predictia Prețului Bitcoin



# CUPRINS

- **Slide 1-3: Introducere**
  - - Titlu: "Predictia Prețului Bitcoin (BTC)"
  - - Introducerea subiectului și obiectivelor prezentării.
- **Slide 4: Contextul pieței Bitcoin**
  - - Bitcoin și a importanței predicției prețului.
  - - Grafice relevante cu evoluția prețului Bitcoin în timp.
- **Slide 5-7: Descrierea setului de date**
  - - Datele folosite în proiect pentru a face predicții.
  - - Numărul de caracteristici și intervalul de timp acoperit.
- **Slide 8-15: Explorarea datelor**
  - - Grafice și vizualizări ale datelor care evidențiază tendințele, corelațiile ale setului de date.
- **Slide 16-20: Metodologia de predicție**
  - - Descrierea algoritmilor pentru a realiza predicția prețului Bitcoin.
  - - Detalii despre modelele de învățare automată utilizate, tehnici de preprocesare a datelor, tunarea modelelor.
  - - Evaluarea performanței modelului (RMSE, MAE, R2)
- **Slide 21-23: Rezultatele predicției**
  - - Prezentarea rezultatelor obținute de modelele.
  - - Compararea predicției cu prețul real al Bitcoin, .
  - - O scurtă descriere despre cum predicțiile pot fi utile pentru investitori, traderi sau alte entități implicate în piața Bitcoin.

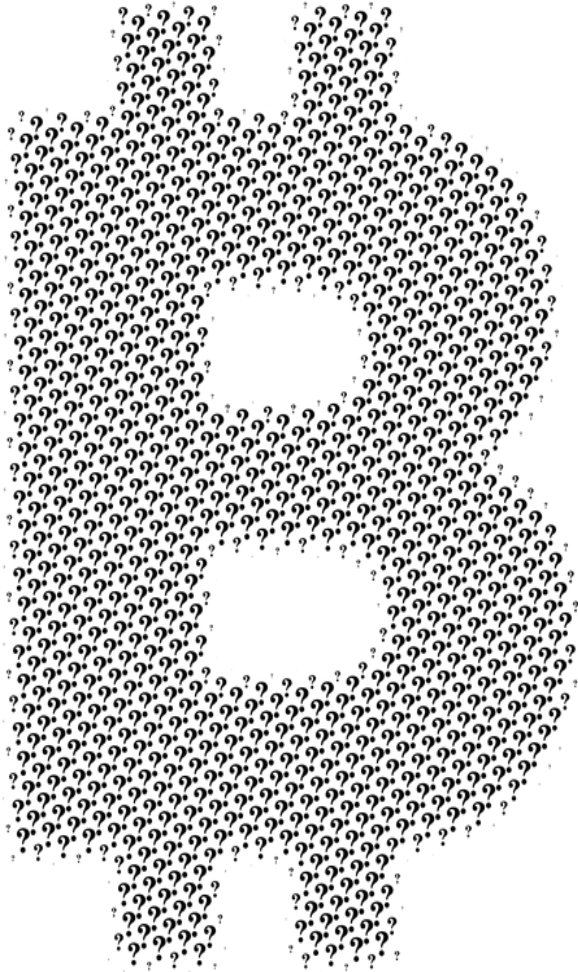
- Subiectul prezentării:
- Acest proiect are menirea de a anticipa și prognoza fluctuațiile prețului Bitcoin, una dintre cele mai discutate și volatile criptomonede din piața actuală. Predicția prețului BTC nu este doar un exercițiu de speculație, ci are implicații profunde în investiții, tehnologie și economie globală.
- Obiectivele prezentării:
- În cadrul acestei prezentări, avem următoarele obiective:
  - Explorarea metodelor de predicție: Vom analiza diverse metode și tehnici folosite pentru a anticipa mișcările de preț ale Bitcoinului, de la modele statistice până la algoritmi de învățare automată și rețele neurale.
  - Prezentarea rezultatelor și interpretarea lor: Vom examina rezultatele obținute din analiza noastră și vom încerca să înțelegem semnificația lor în contextul pieței crypto și al piețelor financiare globale.
  - Evidențierea importanței predicției prețului BTC: Vom sublinia importanța acestui subiect și modul în care predicția prețului Bitcoin poate influența deciziile de investiții, dezvoltarea tehnologică și perspectivele economice viitoare.



## Contextul pieței Bitcoin

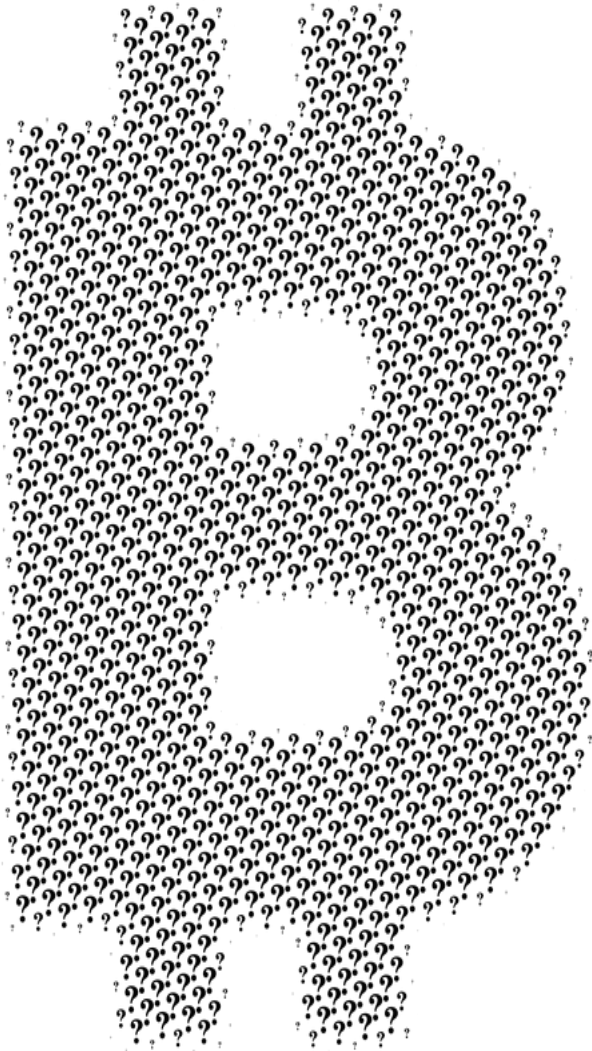
- Bitcoin a fost creat în 2009 de Satoshi Nakamoto , este prima criptomonedă decentralizată și a deschis calea spre o nouă formă de tranzacționare și economie digitală.
- De la valoarea sa inițială Bitcoin a crescut semnificativ în valoare de-a lungul anilor, chiar atingând cifra de 69000\$ și în ciuda volatilității sale a atras atenția ca un instrument de investiții.
- Importanța predicției este binevenită pentru investitori și tranzacționare, stabilitatea pieței, inovație și dezvoltare chiar și adoptare generală ce ar influența încrederea publicului și ar putea contribui la dezvoltare tehnologiei Blockchain.

# DESCRIEREA SETULUI DE DATE



• Proiectul constă din 31 seturi de date și 34 de caracteristici. Datele au fost preluate de pe site-ul oficial [blockchain.com](https://blockchain.com). Intervalul de timp al datelor este cuprins între anii 2009 și ianuarie 2024. Mai jos găsim o mică descriere a fiecărei caracteristici:

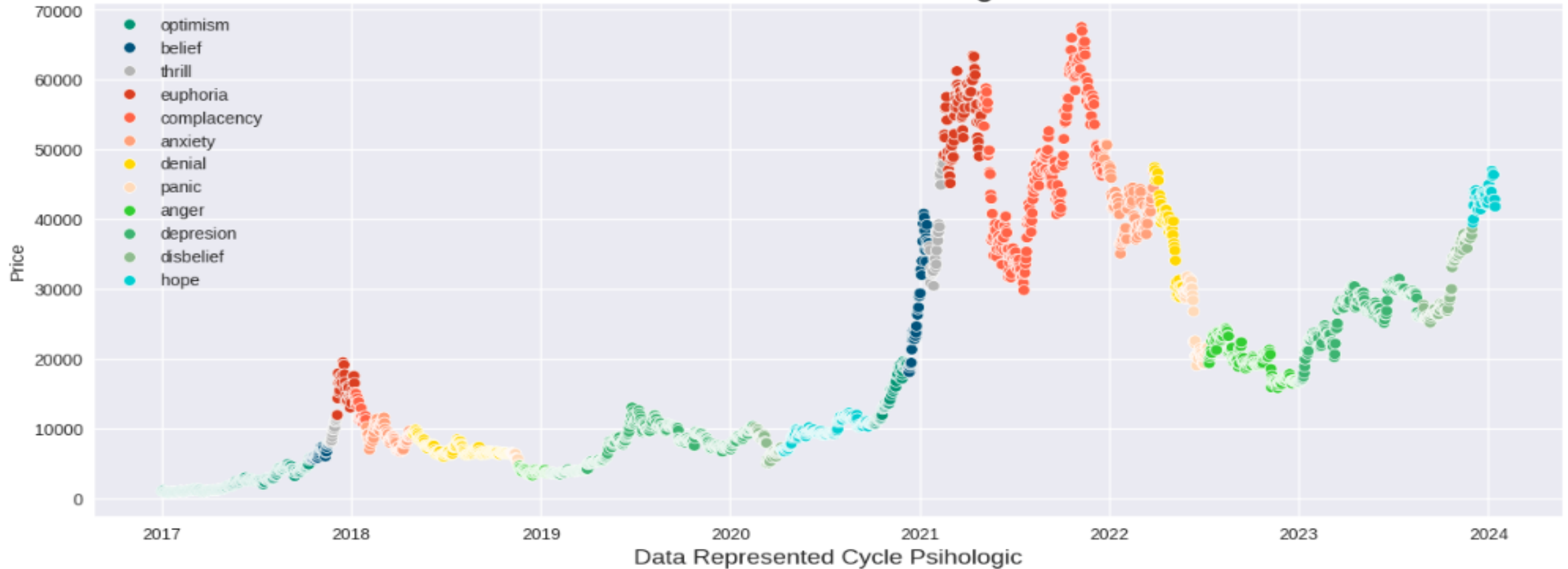
- - 'data': Dată a observației.
- - 'energy\_china\_price': Prețul energiei electrice în China.
- - 'btc\_price': Prețul Bitcoin în USD.
- - 'eur\_usd': Rată de schimb EUR/USD.
- - 'gold\_price': Prețul aurului în USD.
- - 'coal\_china\_price': Prețul cărbunelui în China.
- - 's&p500': Indicele S&P 500.
- - '200w\_moving\_avg': Medie mobilă de 200 de săptămâni.
- - 'avg\_block\_size': Dimensiunea medie a blocurilor Bitcoin.
- - 'avg\_confirmation\_time': Timpul mediu de confirmare pentru tranzacțiile Bitcoin.
- - 'blocks\_size': Dimensiunea blocurilor Bitcoin.
- - 'difficulty\_network': Dificultatea rețelei Bitcoin.
- - 'transaction\_volume\_usd': Volumul total de tranzacții al Bitcoin în USD.
- - 'transaction\_volume\_btc': Volumul total de tranzacții al Bitcoin în BTC.
- - 'transaction\_fees\_usd\_x': Taxele totale de tranzacție ale Bitcoin în USD.
- - 'hash\_rate': Rata de hash a rețelei Bitcoin.



- - 'market\_capital': Capitalizarea de piață a Bitcoin.
- - 'median\_confirmation\_time': Timpul de confirmare median pentru tranzacțiile Bitcoin.
- - 'miners\_revenue': Venitul minierilor de Bitcoin.
- - 'mvrv': Raportul valorii de piață la valoarea realizată a Bitcoin.
- - 'n\_transactions\_per\_block': Numărul de tranzacții per bloc Bitcoin.
- - 'n\_transactions\_total': Numărul total de tranzacții Bitcoin.
- - 'n\_transactions': Numărul total de tranzacții Bitcoin.
- - 'n\_unique\_addresses': Numărul de adrese Bitcoin unice.
- - 'nvt': Raportul valorii rețelei la tranzacții (NVT) Bitcoin.
- - 'nvts': Raportul NVT (valoare rețelei la tranzacții) al Bitcoin, netezit.
- - 'output\_volume': Volumul total de ieșire al tranzacțiilor Bitcoin.
- - 'total\_bitcoins': Numărul total de Bitcoin-uri existente.
- - 'trade\_volume': Volumul de tranzacționare Bitcoin.
- - 'transaction\_fees\_usd\_y': Taxele totale de tranzacție ale Bitcoin în USD.
- - 'transaction\_fees': Taxele totale de tranzacție ale Bitcoin.
- - 'transactions\_per\_second': Numărul de tranzacții Bitcoin pe secundă.
- - 'utxo\_count': Numărul total de UTXO-uri (Unspent Transaction Outputs) Bitcoin.
- - 'psihologi\_cicle': Ciclul psihologic al pieței Bitcoin.

- De asemenea, printre noile caracteristici adăugate în setul de date s-a luat în considerare și prețul energiei electrice din China. Acest aspect este crucial, deoarece China beneficiază de prețuri avantajoase la energia electrică, fapt ce o face un lider mondial în ceea ce privește numărul de mineri de BTC.
- O altă caracteristică esențială pe care nu am putut să o trec cu vederea este prețul cărbunelui din China. Aproximativ 80% dintre centralele electrice din China funcționează pe bază de cărbune, ceea ce indică o strânsă corelație între prețul cărbunelui și prețul energiei electrice. Cu cât prețul cărbunelui este mai scăzut, cu atât energia electrică este mai ieftină și invers.
- De asemenea, cotația S&P500 este un indicator semnificativ al sentimentului general al pieței și al apetitului pentru riscuri. Atunci când investitorii sunt predispuși să asume riscuri, ei vor căuta active precum Bitcoin, astfel încât schimbările în S&P500 ar putea influența prețul Bitcoin.
- Aurul, de-a lungul timpului, a fost considerat un refugiu al activelor. În momentele de incertitudine pe piețele financiare sau de preocupare cu privire la inflație sau stabilitatea valutară, investițiile pot fi orientate către aur sau chiar Bitcoin, ca o protecție împotriva deprecierei valorilor monetare.
- De asemenea, activul EUR-USD poate avea un impact asupra prețului Bitcoin, deoarece o depreciere a dolarului american în raport cu euro ar permite investitorilor din zona euro să acceseze Bitcoin la un preț mai accesibil și invers, ceea ce ar putea influența prețul Bitcoin.

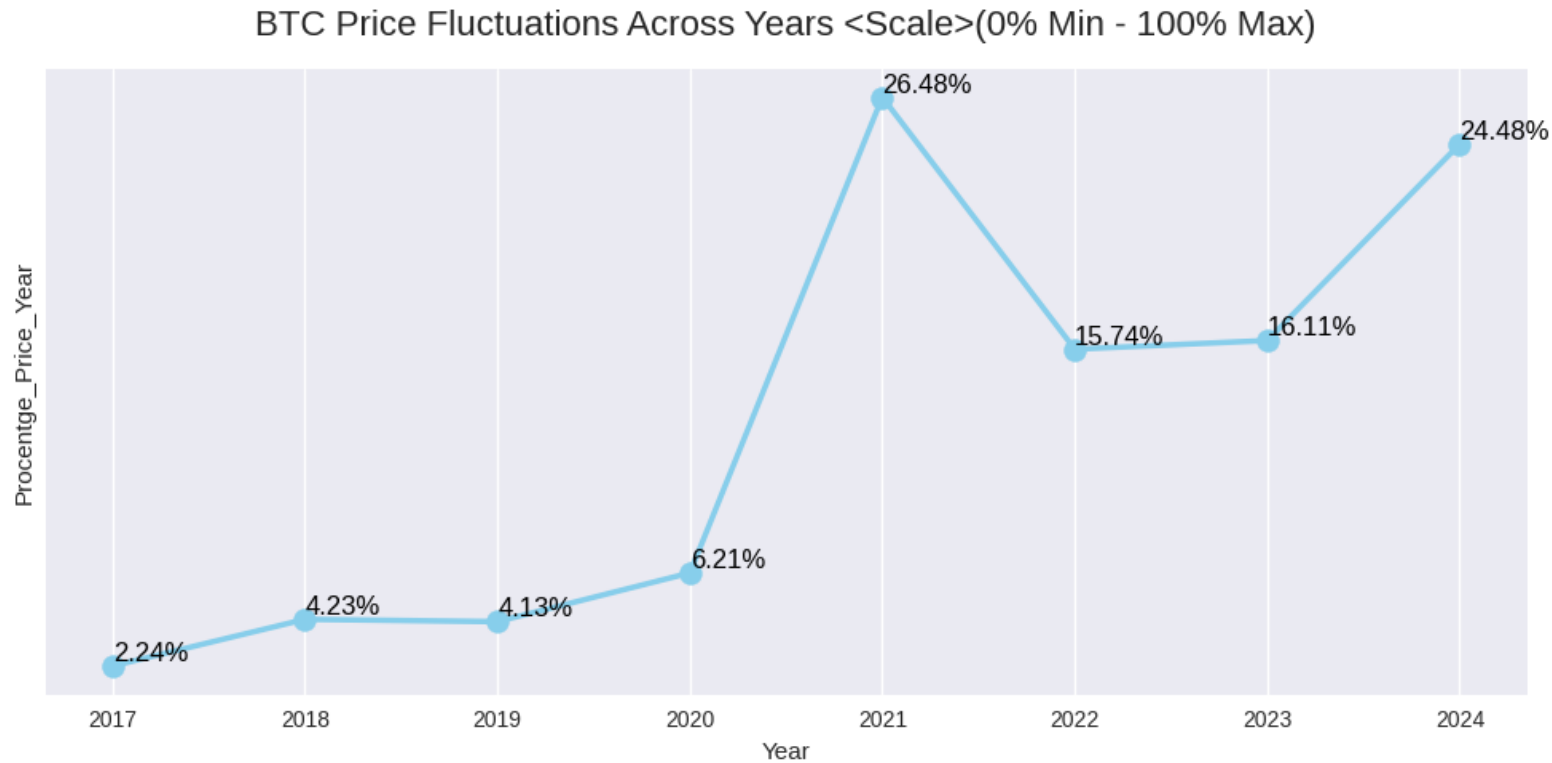
## Price Bitcoin in correlation Psihologi Cicle Human



## EXPLORAREA DATELOR

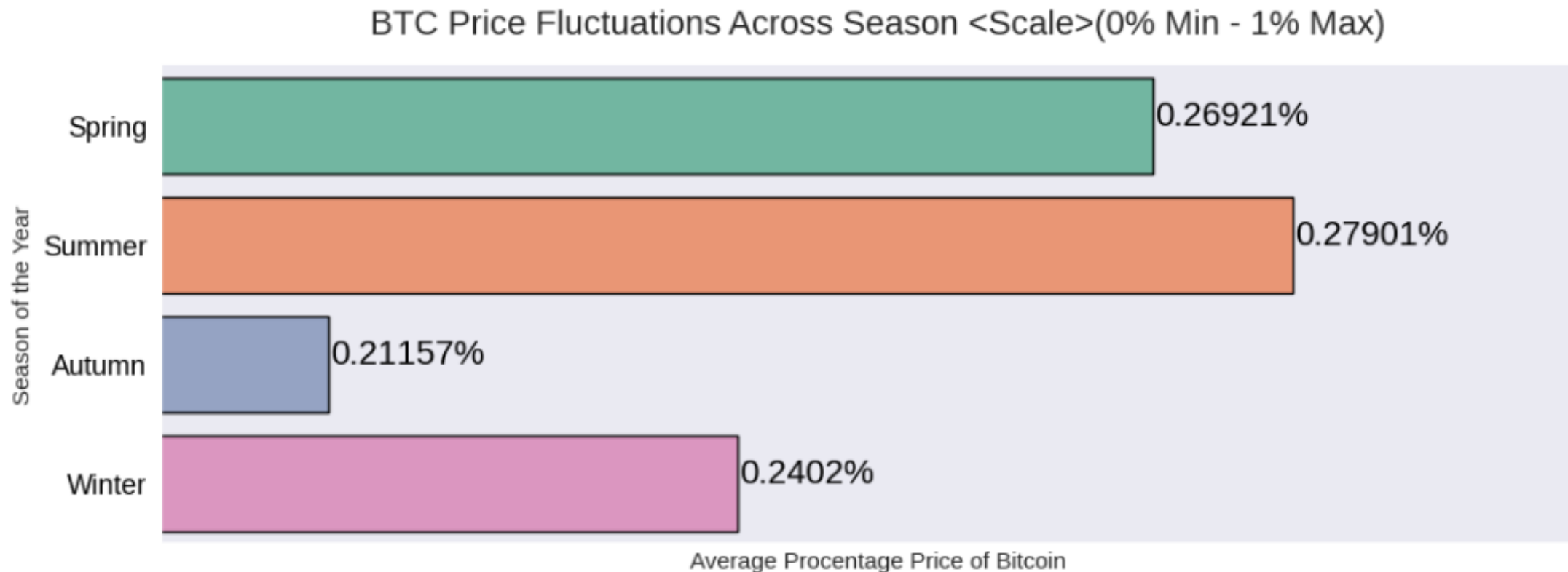
- Acest grafic ilustrează evoluția ciclurilor psihologice umane în timp (2017-2024). Putem observa că investitorii manifestă o stare de speranță în ceea ce privește evoluția prețului.
- Dacă analizăm graficul din anii precedenți, am putea presupune că ne-am putea aștepta ca anul 2024 să fie plin de optimism în ceea ce privește investițiile.

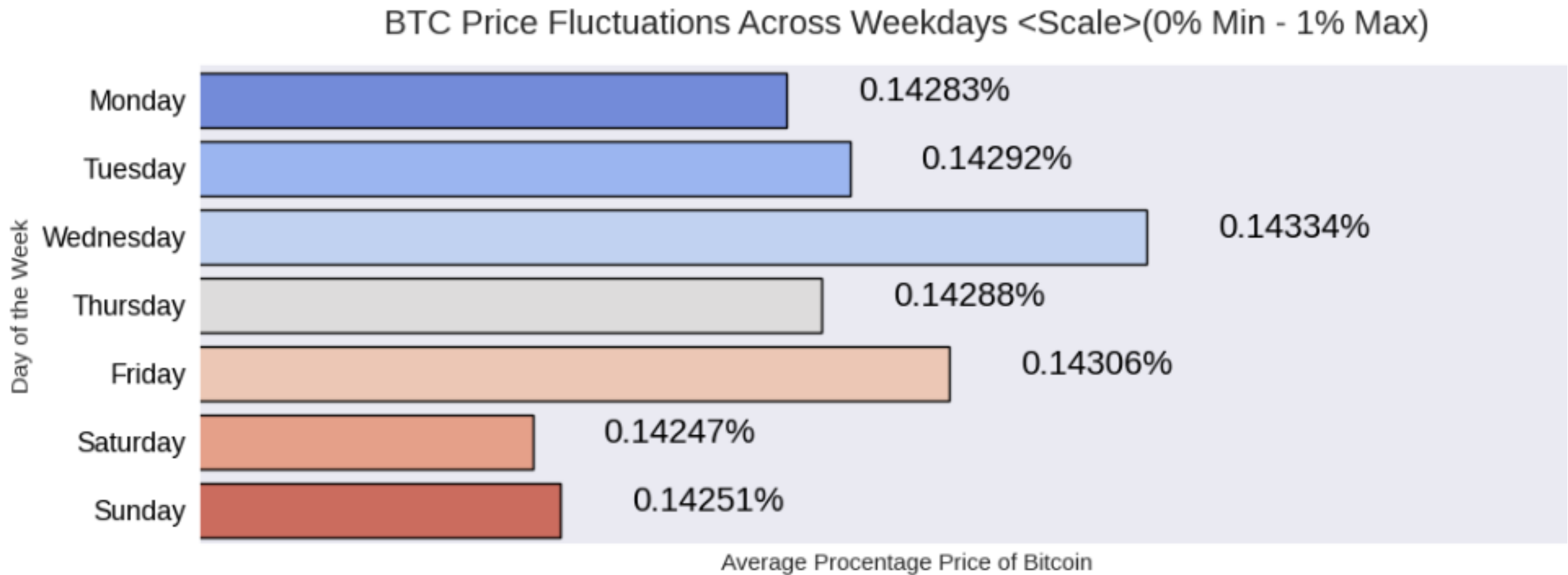




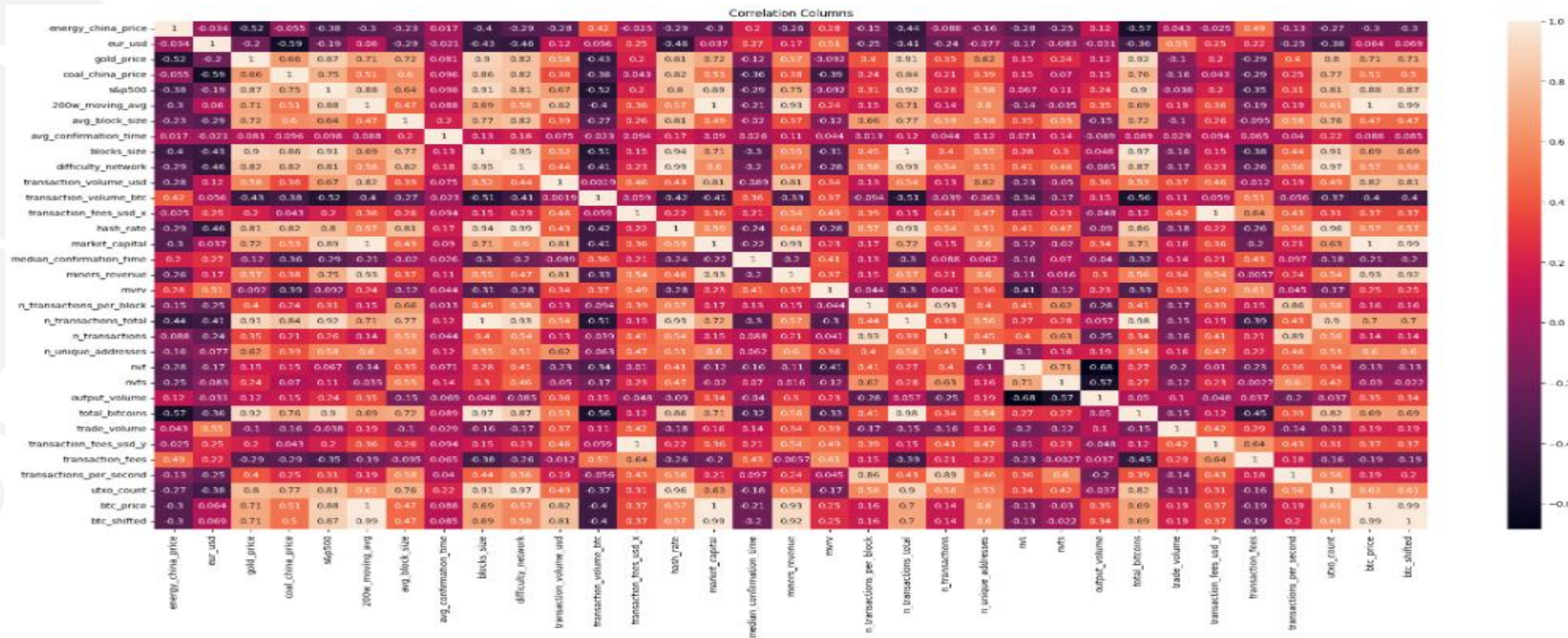
- Acest grafic de tip linear sugerează că Bitcoin se comportă într-un mod neașteptat, indiferent de stadiul în care se află piața, fie în creștere sau în scădere, de la începutul său. În cazul unei comparații, oricum ar indica o tendință pozitivă. Analizând graficul din 2017 până în prezent, putem observa că Bitcoin a înregistrat o creștere mai accentuată începând din anul 2020, acumulând un procent între 16%- 26%, ceea ce indică că pentru investitori Bitcoin reprezintă un refugiu de investiții de încredere.

- Facând o analiză a prețului Bitcoin din anul 2017 până în ianuarie 2024, putem observa că prețul BTC înregistrează o creștere mai semnificativă în perioada verii, cu un procent de 27%, și o creștere mai mică în perioada iernii, cu 21%. Acest lucru sugerează că investitorii sunt mai predispuși să investească primăvara și vara, în timp ce își retrag investițiile în timpul sărbătorilor de iarnă.





Făcând o analiză mai detaliată a datelor, putem observa că BTC înregistrează o creștere generală mai mare în zilele de miercuri și vineri, cu o fluctuație de 14%, în timp ce obține rezultate mai slabe în zilele de weekend. Acest lucru sugerează o corelație strânsă între Bitcoin și bursele monetare internaționale, care sunt închise în zilele de weekend, explicând astfel rezultatele mai slabe acumulate în acele zile.



Acest grafic demonstrează o corelație destul de puternică între anumite coloane care ar putea fi selectate pentru procesare ulterioară. Coloanele care îmi atrag atenția în această etapă sunt: **pret de aur s&p500 coal\_china\_price 200w\_moving\_avg blocuri\_dimensiune rețea\_dificultăți tranzacție\_volum\_usd rata\_hash capital\_piață miners\_revenue n\_transactions\_total adrese\_n\_unice utxo\_count total\_bitcoins** Unele dintre coloane au o corelație destul de puternică, ceea ce ar putea fi un punct negativ pentru modelul meu, crescând șansele de supraadaptare. Aceste coloane au fost incluse inițial în listă, dar în timpul analizei voi decide dacă le păstrez sau le elimin.

Coloanele selectate mai sus au fost trecute prin citevai metode de analiza a datelor si am făcut un scurt rezumat al acestei etape de analiză , iar prin tabelul reprezentat mai jos se poate observa că nu toate coloanele selectate din graficul de corelare prezintă o distribuție uniformă sau o corelație perfectă. Totuși, în urma acestei analize, am decis că coloanele care acumulează cel puțin 2 pozitive vor fi selectate pentru următoarea etapă de preprocesare. Astfel, concluzionăm că din lista noastră, vom șterge 1 coloană și anume: Block\_size.

**Summary Table of Column Selection**

Column	Decomposition	Rolling_Statistic	Dickey_Fuller	Distribution
gold_price	+	+	-	-
s&p500	+	+	-	-
coal_china_price	+	+	-	-
200w_moving_avg	+	+	-	-
blocks_size	+	-	-	-
difficulty_network	+	+	-	-
transaction_volume_usd	+	-	-	+
hash_rate	+	+	-	-
market_capital	+	+	+	-
miners_revenue	+	+	+	-
n_transactions_total	+	-	+	-
n_unique_addresses	+	+	-	-
utxo_count	+	+	+	-
total_bitcoins	+	+	+	-

## **O scurta descriere a metodelor prin care au fost selectate caracteristicile relevante**

**Metoda de Decompozitie** - prin aceasta metoda am identificat tendintele sezoniere si reziduale pentru fiecare coloana. Apoi au fost identificate modelele si comportamentele de trenduri care sunt explicabile , asemenea si corelatiile intre caracteristici cu pretul BTC.

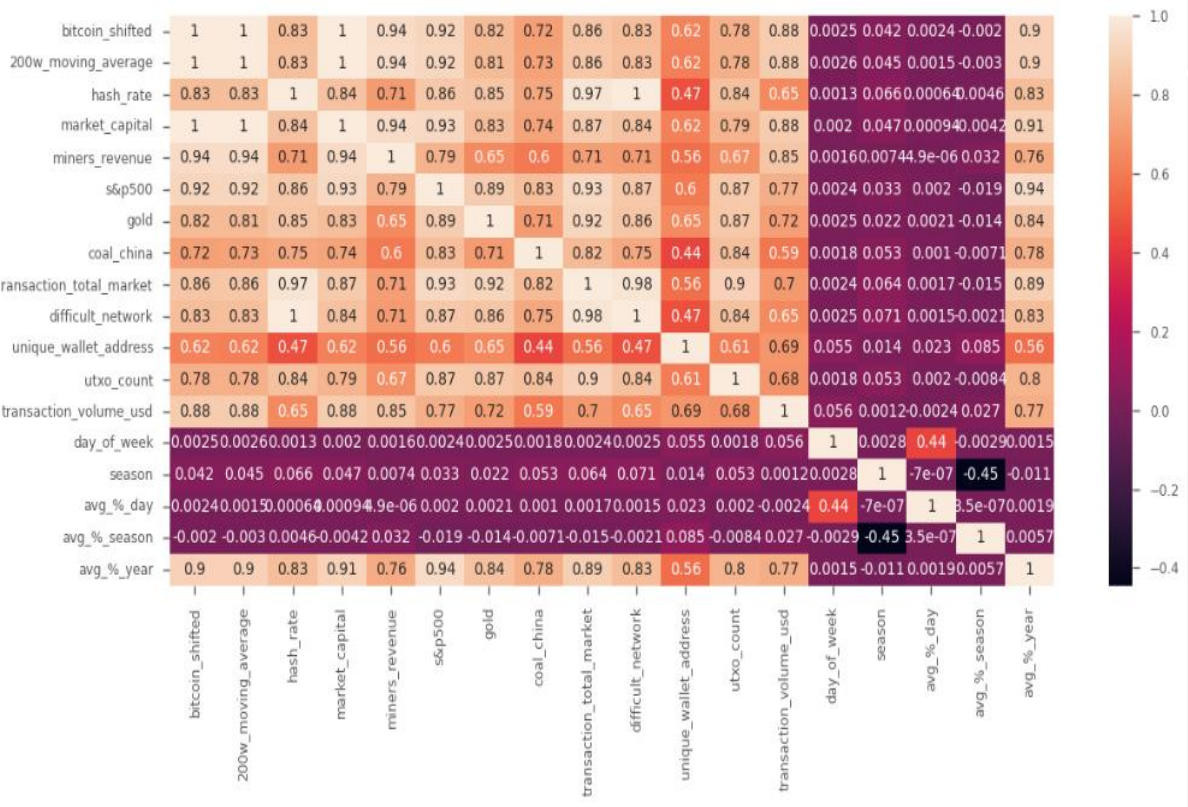
**Metoda Rolling Statistic** - prin aceasta metoda am putut vizualiza trendurile si variabilitatea datelor in functie de timp acest lucru nea permis sa identificam tendințe stabile sau fluctuatii semnificative in timp. Prin plasarea mai multor caracteristici in timp am putut identifica eventualele corelatii si influente reciproce asemeni au fost luate in calcul modelele sezoniere sau ciclice in date. Plotarea acestor date ne permite sa identificam anomaliiile in date.

**Metoda Dickey-Fuller** – prin aceasta metoda au fost selectate caracteristicile cu rezultatul ADF care indica stationaritate adica valoarea P sa fie mica si ADF sub valorile critice, adica coloanele care nu sunt stationare sunt considerate mai putin relevante.

**Metoda Distribution** - vizualizarea distributiilor datelor poate oferi o intelegere mai profunda a variabilitatii si formei distributiei a datelor , caracteristicile cu o distributie simetrica sau normala ar putea fi socotite carateristici potential relevante.



data	2017-01-01	2017-01-02	2017-01-03	2017-01-04	2017-01-05
btc_shifted	902.200989	908.585022	911.198975	902.828003	907.679016
200w_moving_avg	964.84	1000.92	1037.0	1073.08	1109.16
hash_rate	2463610.662209	2400441.15805	2337271.653891	2274102.149732	2210932.645572
market_capital	15638655488.668888	16277109121.505724	15981538928.6884	15685968735.871075	15390398543.053751
miners_revenue	1972367.012584	2108796.052101	2245225.091619	2081796.141733	1918367.191847
s&p500	2253.758341	2255.79421	2257.830078	2270.75	2269.0
gold_price	1150.650024	1155.525024	1160.400024	1163.800049	1179.699951
coal_china_price	2.618332	2.631718	2.645103	2.680371	2.774419
difficulty_network	317688400354.0	317688400354.0	317688400354.0	317688400354.0	317688400354.0
n_unique_addresses	486535.666667	500779.833333	515024.0	512103.25	509182.5
utxo_count	43928455.232639	44036563.472222	44130562.092593	44224560.712963	44318559.333333
transaction_volume_usd	166788994.438506	255752025.269095	344715056.099685	433678086.930274	387648052.087131
n_transactions_total	183451547.0	183961799.0	184264466.25	184567133.5	184869800.75
day_of_week	Sunday	Monday	Tuesday	Wednesday	Thursday
season	Spring	Spring	Spring	Spring	Spring
avg_%_day	0.0389	0.0388	0.0388	0.0389	0.0389
avg_%_season	0.0409	0.0409	0.0409	0.0409	0.0409
avg_%_year	0.0084	0.0084	0.0084	0.0084	0.0084



La aceasta etapa in noul cardru de date vom crea 5 noi caracteristici , acestea sunt : procentul mediu al pretului pe zi , pe sezon si pe an si alte 2 caracteristici vor fi anotimpurile anului si zilele saptaminii.

Coloanele selectate evidențiază un trend nestationar, intentia mea este sa le transform intr-un trend stationare, pentru a face acest lucru vom aplica două metode. Prima metodă este cea logaritmică, iar a doua este descompunerea sezonieră, care ne va furniza toate seriile stationare. Acest lucru ne va ajuta să obținem scoruri mai precise pentru predictivitate și rezultate mai bune ale metricilor.

După aceste preprocesări, obținem un nou set de date cu caracteristici relevante pentru prețul BTC, care prezintă o corelație cuprinsă între 50 și 94%, cu excepția noilor coloane create, care prezintă o corelație destul de slabă. Cu toate acestea, am decis să le păstrăm în cadrul setului de date.

# METODOLOGIA SI PREDICȚIE

În cadrul procesării datelor, setul de date a fost împărțit în:

**X\_train (2450, 17)**

**X\_test (121, 17)**

**y\_train (2450)**

**y\_test (121)**

Ulterior, etichetele X\_train și X\_test au fost scalate folosind metoda standard scale, deoarece avem algoritmi liniari și rețele neuronale care necesită scalare.

Model de antrenare pe care îi vom aplica pe setul de date sunt:

- **LinearRegression()** - Alegerea acestui algoritm se datorează simplității sale și capacității de a oferi o bază de referință pentru evaluarea altor modele mai complexe
- **KNeighborsRegressor()** - A fost ales pentru că este un algoritm non-parametric care poate captura relații complexe între caracteristici și variabila țintă
- **MLPRegressor (rețea neurală)** - acest algoritm poate detecta pattern-uri complexe în datele de intrare și are capacitatea de a modela relații non-liniare între caracteristici
- **Sequential (rețea neurală)** - permite construirea ușoară a unor arhitecturi complexe de rețele neuronale, cum ar fi rețelele neuronale convoluționale sau rețelele neuronale recurente
- **RandomForestRegressor()** - Este un algoritm ensemble care poate gestiona overfitting-ul și poate furniza estimări robuste, fiind adecvat pentru seturi de date cu o mulțime de caracteristici



- **ExtraTreesRegressor()** - este un algoritmu ce îl face mai robust la overfitting
- **XGBRegressor()**- XGBoost este un algoritm de gradient boosting extrem de eficient, cu capacitatea de a gestiona seturi de date mari
- **LGBMRegressor()** - este un alt algoritm de gradient boosting care se concentrează pe eficiență și viteză, fiind potrivit pentru seturi de date mari și complexe
- **AdaBoostRegressor()** - A fost ales pentru capacitatea sa de a îmbunătăți performanța modelului prin combinarea unui număr de modele slabe
- **GradientBoostingRegressor()** - Un alt algoritm de gradient boosting, capabil să modeleze relații complexe și să producă estimări precise
- **DecisionTreeRegressor()** - poate captura interacțiunile complexe între caracteristici fără a necesita prea multă prelucrare a datelor



---

Observam ca din retelele neuronale MLP , am obtinut cel mai bun scor  $r^2$ : 0.85

---

Din modelele lineare LinearRegression are cel mai bun scor  $r^2$ : 0.86 si modelul AdaBoostRegressor cu un scor  $r^2$ : 0.80

---

Restul algoritmilor au inregistrat un scor destul de slab, In etapa urmatoare vom aplica tiuning pe modele, deoarece lucrurile se pot schimba

### AdaBoostRegressor:

- $MSE = 0.008846475648837839$
- $RMSE = 0.09405570503078396$
- $MAE = 0.07268163640788569$
- $r^2 = 0.8054804326559408$

### LinearRegression:

- $MSE = 0.004582210610938365$
- $RMSE = 0.06769202767636943$
- $MAE = 0.05127566923314914$
- $r^2 = 0.8627888769554121$

### MLP:

- $MSE = 0.005025683256722897$
- $RMSE = 0.0708920535513177$
- $MAE = 0.05578814706051883$
- $r^2 = 0.8505938429607646$

# Parametri Tunare și aplicare Penalități

- -- **n\_estimators**: numărul de arbori din modelul de creștere a gradului.
- -- **learning\_rate**: rata de învățare, care controlează cât de mult contribuie fiecare arbore la corectarea erorilor din predicția anterioară.
- -- **max\_depth**: Adâncimea maximă a fiecărui copac din ansamblu.
- -- **subsample**: procentul de eșantionare aleatorie a datelor în construirea fiecărui arbore.
- -- **colsample\_bytree**: procentul de caracteristici (coloane) selectate aleatoriu pentru a construi fiecare arbore.
- -- **reg\_alpha**: Parametrul de regularizare L1.
- -- **reg\_lambda**: Parametrul de regularizare L2.
- -- **gamma**: un parametru de tăiere care controlează cât de mult ar trebui să scadă pierderea pentru a face o nouă divizare pe un nod.
- -- **min\_samples\_split**: numărul minim de mostre necesare pentru a împărți un nod intern.
- -- **min\_samples\_leaf**: numărul minim de mostre necesare pentru a fi pe o frunză de copac.
- -- **max\_features**: numărul maxim de caracteristici de luat în considerare atunci când căutați cea mai bună împărțire.

## MLP

- alpha=0.001
- beta\_1=0.9
- beta\_2=0.999

## KERAS

- kernel\_regularizer=regularizers.l1\_l2
- bias\_regularizer=regularizers.l1\_l2

In urma tunării algoritmilor de regresie lucrurile nu s-au îmbunătățit dramatic dar totuși aplicarea penalității algoritmilor neuronali putem vizualiza ca aceste manipulări ne-au ajutat să îmbunătățim metricile ...

Pentru MLP  $r^2$ : 0.88

Pentru KERAS  $r^2$ : 0.92

Totuși ca rezultat valoros îl putem socoti algoritmul neuronal SEQUENTIAL biblioteca KERAS cu un scor  $r^2$ : 0.92 și MAE: 0.0373 ceea ce avem ca dovadă o bună performanță a modelului între predicție și valorile reale a setului de date.

For the MLP:

- MSE = 0.0038460050884882822
- RMSE = 0.062016167960365645
- MAE = 0.045336229539834315
- $r^2$  = 0.8856639364496938

And for Sequential, the best score:

- MSE = 0.0025708550747888825
- RMSE = 0.050703600215259687
- MAE = 0.0373461048751258
- $r^2$  = 0.9235722672105913

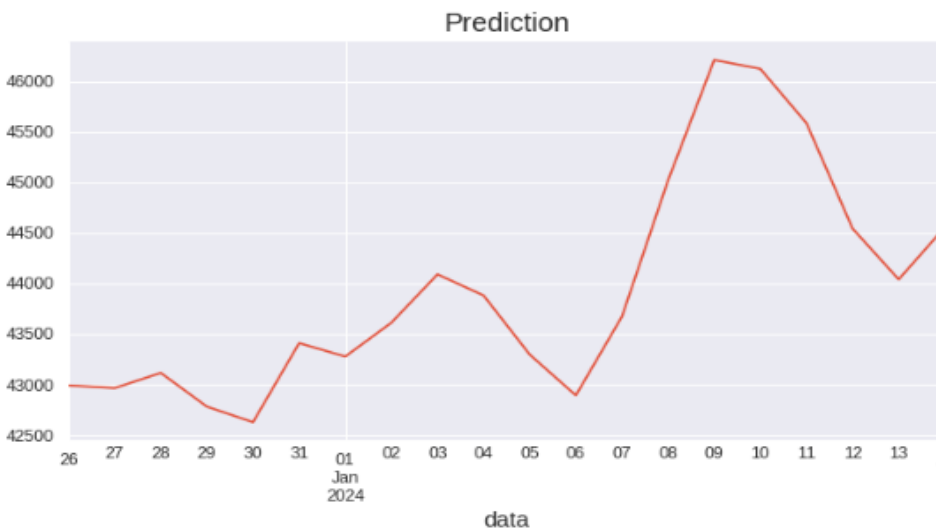
Layering Charts: Real vs. Predicted Bitcoin Prices Compared to Human Psychology



## Vizualizăm Rezultatul Predicției

- În acest grafic de tip linie, putem observa că investitorii se află într-o stare psihologică de speranță. De asemenea, putem remarca că prețul Bitcoin prezintă o tendință de creștere, ceea ce ar stimula încrederea și ar deschide noi oportunități de investiții pentru traderi și investitori.

- Prin aceste două grafice de tip linie, putem observa că predicția făcută de rețeaua neuronală prezintă o mică deviere față de prețul real al Bitcoin. Acest lucru este confirmat și de tabelul care prezintă două coloane: una cu prețul real și alta cu prețul prezis. De asemenea, putem observa că modelul a realizat predicții destul de precise în unele zile, având o abatere de 1.08% față de prețul real, în timp ce în alte zile observăm o abatere de până la 6.19%, ceea ce reprezintă o eroare destul de semnificativă.
- Pentru a îmbunătăți predicțiile modelului, este necesară utilizarea unor date mai noi și mai precise, precum și o examinare mai detaliată și ajustarea meticuloasă a algoritmilor pentru a îmbunătăți precizia predicției în analiza prețului



	Prediction	btc_price
data		
2023-12-26	42988.085938	42520.402344
2023-12-27	42961.609375	43442.855469
2023-12-28	43112.484375	42627.855469
2023-12-29	42777.921875	42099.402344
2023-12-30	42624.070312	42156.902344
2023-12-31	43407.421875	42265.187500
2024-01-01	43274.289062	44167.332031
2024-01-02	43609.621094	44957.968750
2024-01-03	44088.550781	42848.175781
2024-01-04	43879.027344	44179.921875
2024-01-05	43293.609375	44162.691406
2024-01-06	42889.234375	43989.195312
2024-01-07	43670.839844	43943.097656
2024-01-08	45020.761719	46970.503906
2024-01-09	46212.648438	46139.730469
2024-01-10	46124.457031	46627.777344
2024-01-11	45586.738281	46368.585938
2024-01-12	44542.359375	42853.167969
2024-01-13	44038.503906	42842.382812
2024-01-14	44555.449219	41796.269531

Pentru investitori, predicțiile pot oferi o orientare valoroasă în luarea deciziilor de investiții, ajutându-i să identifice oportunități de tranzacționare și să gestioneze riscurile asociate cu volatilitatea prețului Bitcoin.



Pentru traderi, modelele pot servi drept instrumente esențiale în optimizarea strategiilor lor de tranzacționare, permitându-le să efectueze tranzacții mai eficiente și mai profitabile.



Alte entități implicate în piața Bitcoin, cum ar fi platformele de tranzacționare sau fondurile de investiții, pot beneficia de predicții pentru a planifica și gestiona riscurile asociate cu expunerea lor la acest activ volatil.



În concluzie, modelele de predicție reprezintă instrumente esențiale pentru gestionarea portofoliilor și pentru optimizarea performanței investiționale în cadrul peisajului complex al criptomonedelor