**Identify Block of Pending Transactions**

**1. Import Required Libraries**

The following libraries were imported for data processing and analysis:

* numpy
* pandas
* matplotlib
* seaborn
* os
* json

**2. Load Dataset**

* Data was loaded from Google Drive.
* Two DataFrames were created:
  + **Pending Transactions:** Unconfirmed transactions still in the queue.
  + **Mined Transactions:** Transactions successfully added to a block.

**3. Exploratory Data Analysis (EDA)**

**a. Sample Head**

* The dataset was examined by displaying the first few rows.

**b. Featureset**

**Pending Transactions Features:**

['type', 'chainId', 'nonce', 'gas', 'maxFeePerGas', 'maxPriorityFeePerGas', 'to', 'value', 'accessList', 'input', 'r', 's', 'yParity', 'v', 'hash', 'blockHash', 'blockNumber', 'transactionIndex', 'from', 'gasPrice', 'blobVersionedHashes', 'maxFeePerBlobGas']

**Mined Transactions Features:**

['hash', 'type', 'accessList', 'blockHash', 'blockNumber', 'transactionIndex', 'confirmations', 'from', 'to', 'nonce', 'data', 'r', 's', 'v', 'creates', 'chainId', 'gasPrice.type', 'gasPrice.hex', 'maxPriorityFeePerGas.type', 'maxPriorityFeePerGas.hex', 'maxFeePerGas.type', 'maxFeePerGas.hex', 'gasLimit.type', 'gasLimit.hex', 'value.type', 'value.hex']

**c. Shape**

* **Pending Transactions:** 431 rows × 22 features.
* **Mined Transactions:** 500 rows × 26 features.

**d. Missing Data**

**Pending Transactions Missing Data:**

* **Completely Missing (100%):** transactionIndex, blockNumber, blockHash
* **Highly Missing (99.5%+):** maxFeePerBlobGas, blobVersionedHashes, accessList, yParity
* **Other Missing Data:** maxFeePerGas (95), maxPriorityFeePerGas (95), chainId (2)

**Mined Transactions Missing Data:**

* **Completely Missing (100%):** creates
* **Highly Missing (95%+):** accessList
* **Other Missing Data:** maxFeePerGas.hex (59), maxPriorityFeePerGas.hex (59), maxFeePerGas.type (59), maxPriorityFeePerGas.type (59)

**4. Data Preprocessing**

* **Handled missing data** using the mode method for:
  + Pending transactions: chainId, maxFeePerGas, maxPriorityFeePerGas
  + Mined transactions: maxFeePerGas.type, maxFeePerGas.hex, maxPriorityFeePerGas.type, maxPriorityFeePerGas.hex

**5. Feature Analysis**

* Analyzed and visualized features of pending and mined transactions.

**6. Data Transformation**

* Converted hex values to decimal for:
  + type, chainId, nonce, v

**7. Feature Renaming**

* Mined transactions:
  + 'maxFeePerGas.hex' → 'maxFeePerGas'
  + 'maxPriorityFeePerGas.hex' → 'maxPriorityFeePerGas'
  + 'value.hex' → 'value'
  + 'data' → 'input'
  + 'gasPrice.hex' → 'gasPrice'
  + 'gasLimit.hex' → 'gasLimit'
* Pending transactions:
  + 'gas' → 'gasLimit'

**8. Dropping High-Missing Data Columns**

**Pending Transactions:**

Dropped:

* transactionIndex, blockHash, blockNumber (100% missing)
* maxFeePerBlobGas, blobVersionedHashes (99.5% missing)
* accessList (99.4% missing)
* Identity-related columns: chainId, yParity, transactionIndex

**Mined Transactions:**

Dropped:

* accessList (96% missing)
* Identity and type-related columns: ['accessList', 'creates', 'gasPrice.type', 'maxPriorityFeePerGas.type', 'maxFeePerGas.type', 'gasLimit.type', 'value.type']

**9. Data Types**

Verified and converted data types where necessary:

| **Feature** | **Type** |
| --- | --- |
| Type | int64 |
| chainId | int64 |
| Nonce | int64 |
| gasLimit | object |
| maxFeePerGas | object |
| maxPriorityFeePerGas | object |
| To | object |
| Value | object |
| Input | object |
| R | object |
| S | object |
| V | int64 |
| Hash | object |
| From | object |
| gasPrice | object |

**10. Feature Correlation Analysis**

**Pending Transactions:**

Correlation Matrix:

* Strongest correlations:
  + maxFeePerGas & gasPrice: **0.90**
  + type & v: **-0.97**
  + maxFeePerGas & type: **-0.30**

**Mined Transactions:**

Correlation Matrix:

* Strongest correlations:
  + gasPrice & maxPriorityFeePerGas: **0.63**
  + maxFeePerGas & v: **0.31**
  + type & v: **-0.97**

**Heatmaps were plotted for visualization.**

**11. Dropped Low-Correlation and Handled Multi-Correlation Features**

* **Pending Transactions:** Dropped 'type', 'chainId'
* **Mined Transactions:** Dropped 'type', 'confirmations', 'chainId', 'transactionIndex'

**12. BlockNumber Exploration in Mined Transactions**

* Transactions were evenly distributed across five blocks:
  + Block 21719048 → 100 transactions
  + Block 21719049 → 100 transactions
  + Block 21719045 → 100 transactions
  + Block 21719047 → 100 transactions
  + Block 21719046 → 100 transactions

**13. Final Insights**

* **Pending transactions have significant missing data**, especially in block-related columns.
* **Mined transactions are well-distributed across multiple blocks.**
* **Gas price and fee metrics are key differentiators** in both pending and mined transactions.
* **Feature engineering and data cleaning steps improved dataset quality**, making it more suitable for further analysis.