Financial (Stock) Data Collection and Storage

By Steve Champion

Overview

- Project in two parts:
  - Data collection and storage
  - Data cleaning and analysis

The Historic Options Database

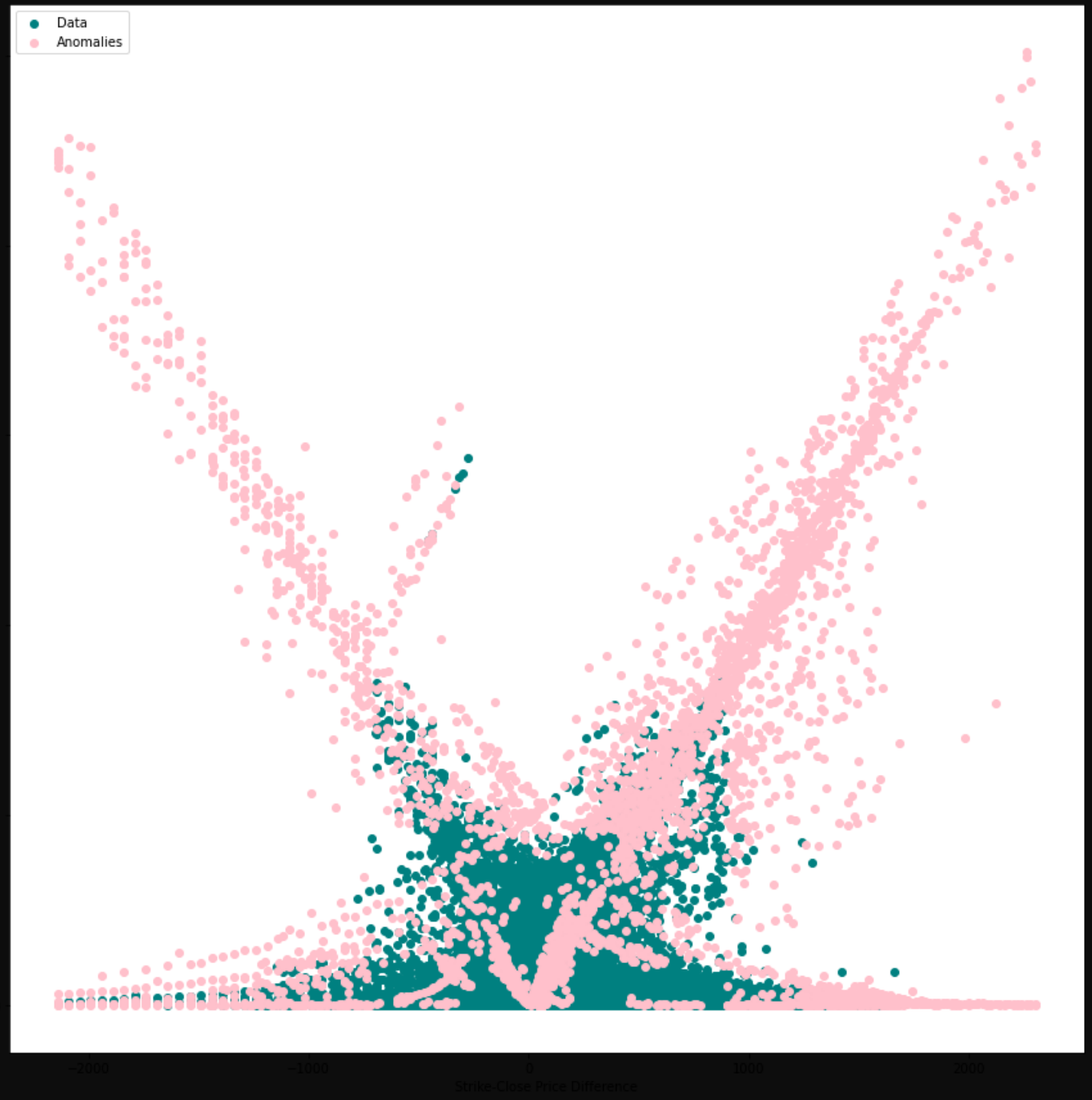
- Historic options data not widely available
- Built framework for historic options collection and storage
- In collection process (along with historic stock data)
  - MySQL Database
  - Already contains useful data

The Data

- Missing volume data
- Placeholder values for implied volatility
- GAIN applied to both
  - When complete will be able to generate reasonable values for missing data

Anomaly Detection

- Used an “Isolation Forest” model
- Found anomalous data
  - Contained obvious anomalies as well as subtle
    - High price -> anomalous
    - Extreme strike-stock price difference -> anomalous
    - Many anomalies near “normal” data
  - To be further investigated
  - Potential for arbitrage trading



Future Plans

- Rework the GAIN framework
  - A successful GAIN will have ability to generate reasonable values for missing data
- Investigate Anomalous data at a deeper level
  - Search for inefficiencies with purpose of generating arbitrage trading strategies
  - Learn more about underlying mechanics of options

Conclusion

- Historical options data is rich with information
  - Not yet ubiquitously collected or studied (on retail side
    - Evidenced by lack of availability
- MySQL is a highly customizable and scalable database management system
  - Useful for automated collection and storage of stock and stock derivatives data

Email any questions to
stevenchamp11@yahoo.com