Classified songs based on genre and popularity using lyrics only by representing lyrics using a distributed representation f sentences and documents. Implemented both traditional machine learning and deep learning models using python.

Achieved an F1 score 8% higher than existing models.

Statistically captured and analyzed social biases on songs lyrics.

* Develop machine learning models using tools like Python, R, Presto & SQL to address problems in areas such as Construction Cost, Facility Design, Facility Operations, Space Planning, Culinary, and/or Transportation
* Drive impactful and thoughtful solutions by collaborating with members of the Analytics team and other technical partners on project planning, design, and execution
* Manage workload with minimal supervision and deliver according to project schedule; proactively provide visibility on progress by leveraging internal project management and development tools, escalating blockers with team and project leads as necessary
* Apply development best practices by leveraging version control tools (e.g., Git, Mercurial); participating in peer code reviews; and incorporating data and model validation steps as part of your standard workflow to produce reproducible, robust results
* Communicate data findings and model results to stakeholders and other team members with clear summaries and recommendations to inform next steps
* Identify and prioritize improvements and enhancements to data needed to support future modeling efforts and coordinate with data engineering partners on data model design & build

**Minimum Qualifications**

* Degree in a quantitative field which emphasizes scientific thinking (e.g., Statistics, Mathematics, Econometrics, Computer Science) or equivalent experience in a data science or related role
* At least 4 years of professional experience in a data science or related role
* Experience identifying appropriate statistical and/or algorithmic approaches to address business problems
* Experience using Python or R to build statistical and/or machine learning models such as time series forecasting, clustering, regression, or classification using at least two of the following algorithms: ARIMA, Linear Regression, decision trees, random forest, XGBoost, logistic regression, K-Means, DBSCAN
* Experience writing SQL queries to extract and prepare data for modeling and analyses
* Experience applying modeling fundamentals and best practices when building descriptive and predictive models (e.g., feature selection, hyperparameter tuning, Bias-Variance Tradeoff)