VBigQuery ML supports the following types of models:

* [Linear regression](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create) for forecasting; for example, the sales of an item on a given day. Labels are real-valued (they cannot be +/- infinity or NaN).
* [Binary logistic regression](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create) for classification; for example, determining whether a customer will make a purchase. Labels must only have two possible values.
* [Multiclass logistic regression](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create) for classification. These models can be used to predict multiple possible values such as whether an input is "low-value," "medium-value," or "high-value." Labels can have up to 50 unique values. In BigQuery ML, multiclass logistic regression training uses a [multinomial classifier](https://en.wikipedia.org/wiki/Multinomial_logistic_regression) with a [cross-entropy loss function](https://developers.google.com/machine-learning/glossary/#cross-entropy).
* [K-means clustering](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create) for data segmentation; for example, identifying customer segments. K-means is an unsupervised learning technique, so model training does not require labels nor split data for training or evaluation.
* [Matrix Factorization](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-matrix-factorization) for creating product recommendation systems. You can create product recommendations using historical customer behavior, transactions, and product ratings and then use those recommendations for personalized customer experiences.
* [Time series](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-time-series) for performing time-series forecasts. You can use this feature to create millions of time series models and use them for forecasting. The model automatically handles anomalies, seasonality, and holidays.
* [Boosted Tree](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-boosted-tree) for creating [XGBoost](https://xgboost.readthedocs.io/en/latest/) based classification and regression models.
* [Deep Neural Network (DNN)](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-dnn-models) for creating TensorFlow based Deep Neural Networks for [classification](https://www.tensorflow.org/api_docs/python/tf/estimator/DNNClassifier) and [regression](https://www.tensorflow.org/api_docs/python/tf/estimator/DNNRegressor) models.
* [AutoML Tables](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-automl) to create best-in-class models without feature engineering or model selection. [AutoML Tables](https://cloud.google.com/automl-tables) searches through a variety of model architectures to decide the best model.
* [TensorFlow model importing](https://cloud.google.com/bigquery-ml/docs/reference/standard-sql/bigqueryml-syntax-create-tensorflow). This feature lets you create BigQuery ML models from previously trained TensorFlow models, then perform prediction in BigQuery ML.

DB SET – BigQuery – worldpop

SELECT lower(concat(title, " ", regexp\_replace(body, r"<[^>]\*>", ""))),

    replace(tags, "|", ",") as tags

FROM `bigquery-public-data.stackoverflow.posts\_questions`

WHERE regexp\_contains(tags, r"(?:tensorflow|keras|matplotlib|pandas|scikit-learn)")

SELECT

regexp\_replace(f0\_, r"(?:tensorflow|tf|keras|matplotlib|plt|pandas|pd|sklearn|scikit-learn)", "bodybuilding")

FROM `stackoverflowmltest.soDataSet.soPostsQuery`

gs://rostmlsb/stackoverflowMLdata.csv