ANITA B.ORG 2023 GRACE HOPPER CELEBRATION | THE WAY FORWARD

Friend up Your Cash App Game

Anusha Jamkhandi
Julie Quintero
Rashmi Raghunandan

About Us

Julie Quintero
Machine Learning Engineer
Cash App
mariajuliana@squareup.com

Rashmi Raghunandan
Machine Learning Engineer
Cash App
rashmir@squareup.com

Anusha Jamkhandi
Machine Learning Engineer
Cash App
ajamkhandi@squareup.com

Agenda

- Building a ML Pipeline with Prefect
- Hands-On: Setup Prefect & GCP
- Introduction to BigQuery
- Hands-On: Upload data to BigQuery via Prefect
- Data Exploration Techniques
- Hands-On: Data Exploration Challenge
- Feature Encoding for ML
- Hands-On: Model Exploration & Embeddings

Disclaimer: Imaginary Data Alert!

The data you're about to see is purely a product of our collective imagination. It's the stuff of dreams, the figment of our data wizards' creativity.

Github Repo:
https://tinyurl.com/friend-up-your-cash-app-game

Collab Notebook:
https://tinyurl.com/friend-up-your-cash-game-nb

Building an ML Pipeline

Understanding Data
Visualizing Insights
Engineering features

Algorithm Selection
Training and Validation
Hyperparameter Tuning

Deployment
Input Processing
Output Generation

EXPLORATION

TRAINING

INFERENCE

Building an ML Pipeline

Understanding Data
Visualizing Insights
Analyzing features

Algorithm Selection
Training and Validation
Hyperparameter Tuning

Deployment
Input Processing
Output
Generation

EXPLORATION

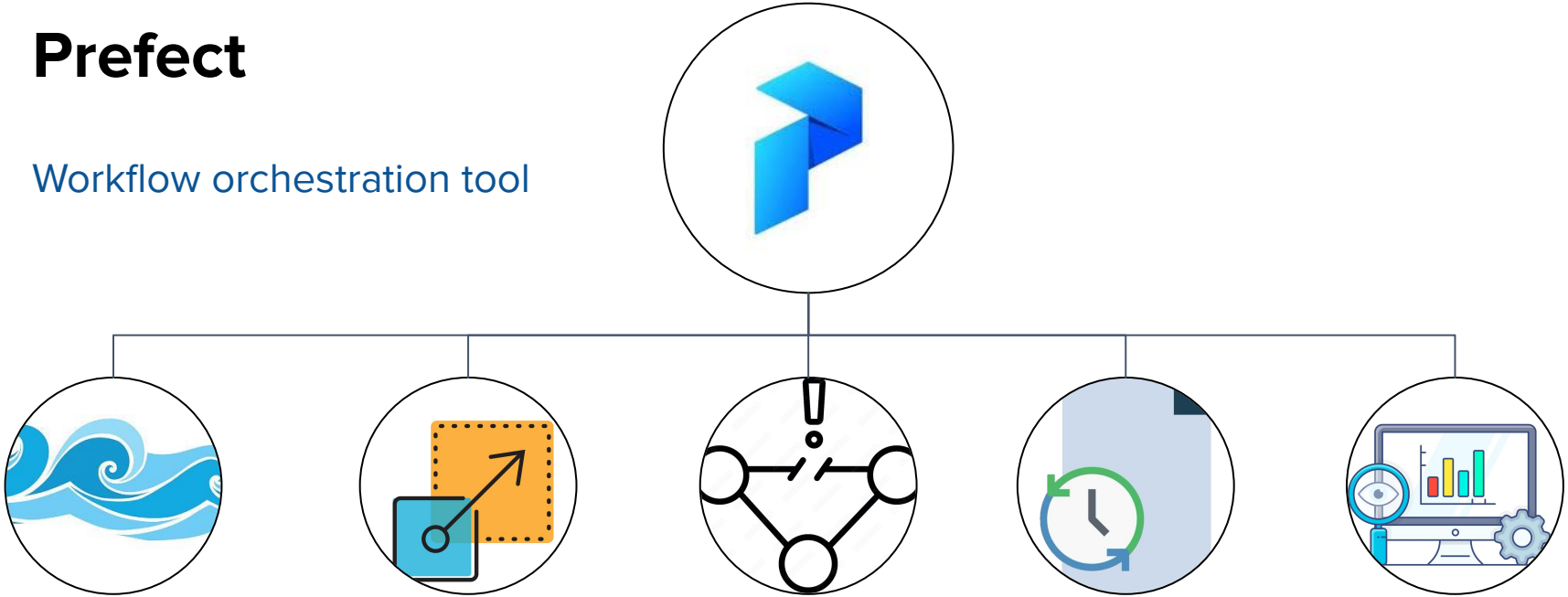
Error

TRAINING

INFERENCE

Prefect

Workflow orchestration tool



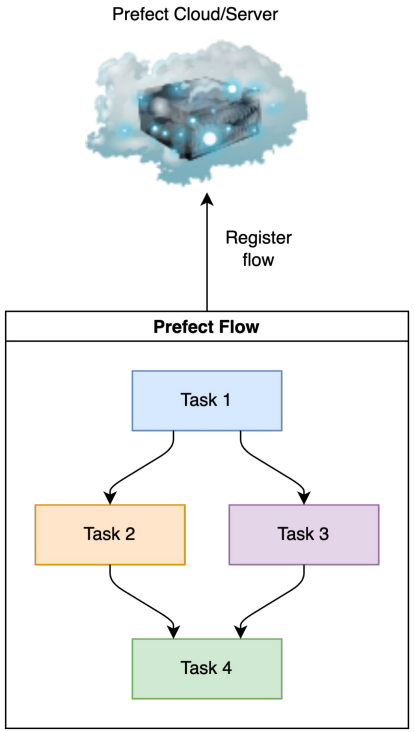
Prefect Architecture

Task:

- Represents a single unit of work
- Can be a python function or a callable

Flow:

- Collection of tasks in a Directed Acyclic Graph (DAG)
- Represents your ML pipeline
- Monitors workflow runs



ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Setup Prefect

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Setup GCP

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Data Infrastructure and Exploration

Setup GCP Infrastructure
Upload Data into BigQuery
Explore Data

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

BigQuery

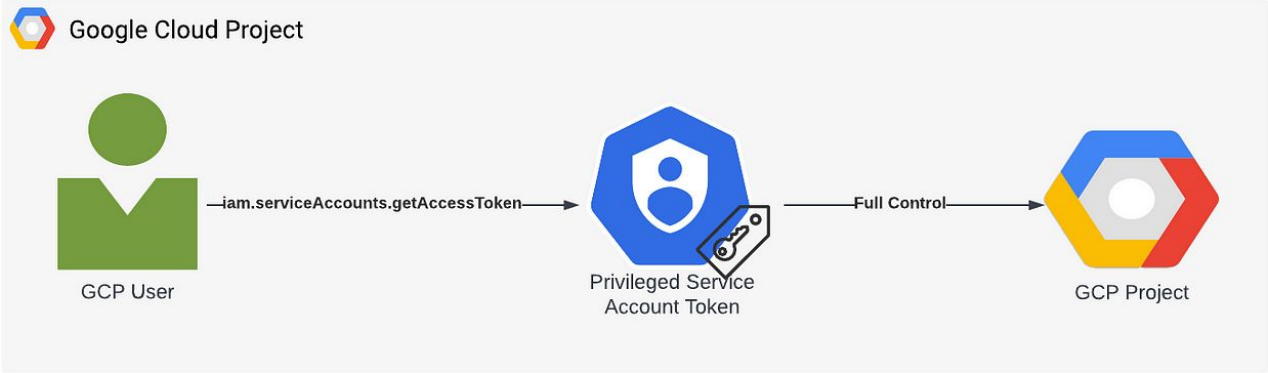
- A serverless, fully-managed data warehouse by Google Cloud.
- Designed for speed and scalability, analyzing massive datasets effortlessly.

- **Dataset:**
  - A container that holds tables, views, and other dataset-specific metadata,
  - Provides a structured way to organize and manage your data within the platform.

- **BigQuery Table:**
  - A structured representation of data in BigQuery.
  - Organized in rows and columns, with defined schema.
  - Supports SQL-like queries and joins for data exploration.

Service Accounts

- A Service Account is a Google Cloud identity
- Used for authenticating applications and services
- Allows controlled access to Google Cloud resources



What is the Purpose of the Key?

- Confidential piece of information used to securely generate digital signatures and authenticate API requests.
- It ensures data integrity and secure communication between applications and GCP services.

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Create Table and upload data using Prefect

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

Data Exploration

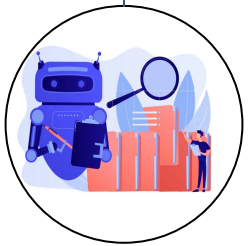
- Understand patterns, trends, relationships
- Assess data quality (missing, outliers)
- Select relevant features
- Validate assumptions
- Support informed decisions

Data Exploration Challenge https://tinyurl.com/dataexplorationchallenge

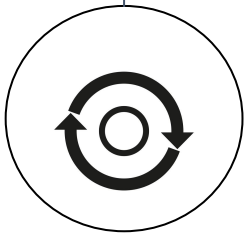
- **Basic Stats:**
  - How many rows are in the dataset?
  - How many distinct rows are there?.
- **Example Cash App Usage:**
  - Find the number of users who used Cash App for less than 1 year.
  - Find the number of users who used Cash App for more than 8 years.
- **Transaction Amount:**
  - Calculate the 99th percentile of the transaction amount.
- **Most Interacted Users:**
  - Determine the count of mutual interactions among the most interacted users
- **Feature Correlation:**
  - Discover if any features (columns) are correlated with each other.

Infrastructure as Code (IaaC)

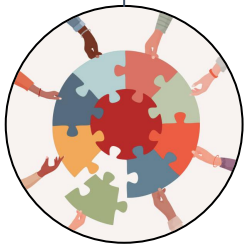
Managing and Provisioning Infrastructure using Code



Automation



Consistency



Collaboration

Reliability

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Explore Model

Encoding features
Distance Metrics
Embeddings

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

Encoding Our features

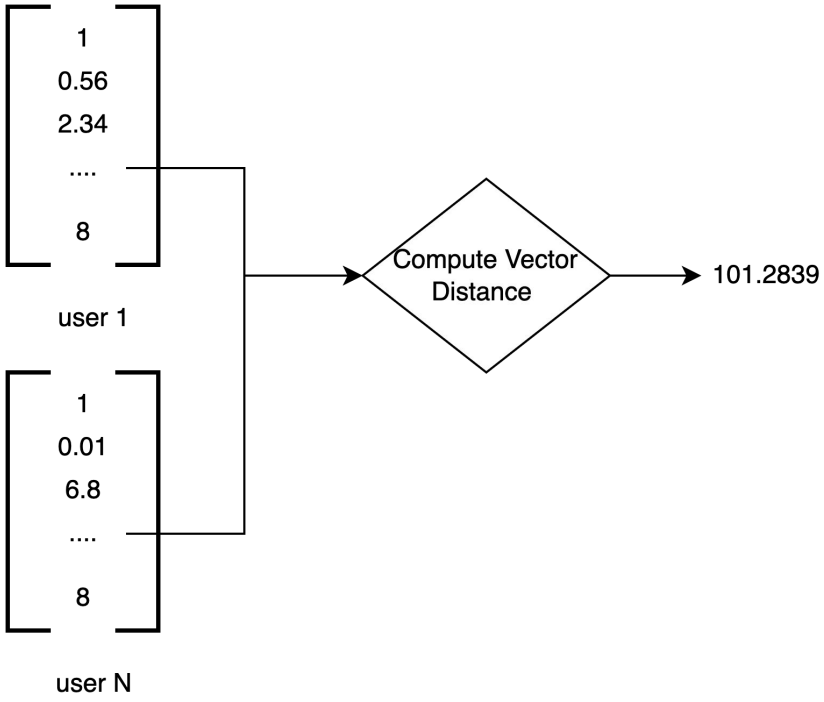
Categorical Encoding

user_occupation	Encoded value
Accountant	0
Lawyer	1
Engineer	2
. . .	. . .
Doctor	N

Binary Encoding

cash_boost_used	Encoded value
Yes	1
No	0

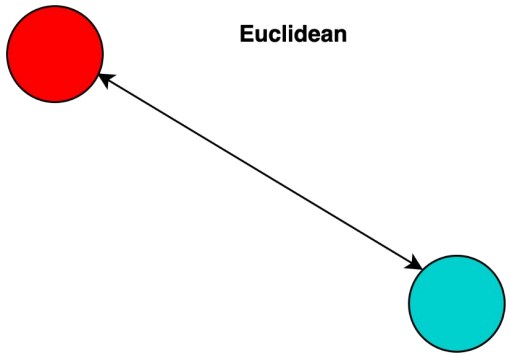
Intro to Vector Similarity



Vector Distance Metrics

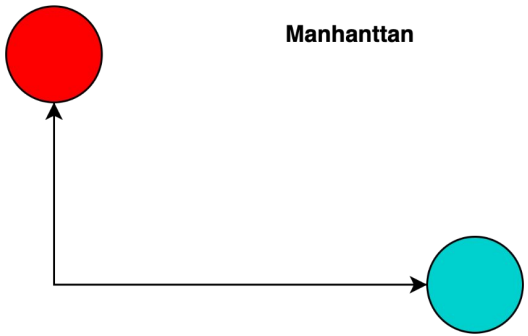
Euclidean Distance

is the straight-line distance between two points in a space. It's commonly used when the data points have continuous numeric attributes.



Manhattan Distance

measures the distance between two points by summing the absolute differences along each dimension.



Let’s compute these distances!

Compute the each of these vector distances between user 1 and the rest of our users in the cash app dataset !

Let’s sort the results and return the top 5 most “similar” users according to each metric

Notebook Code:
https://tinyurl.com/friend-up-your-cash-game-nb

Solution:
https://tinyurl.com/modelling-solution

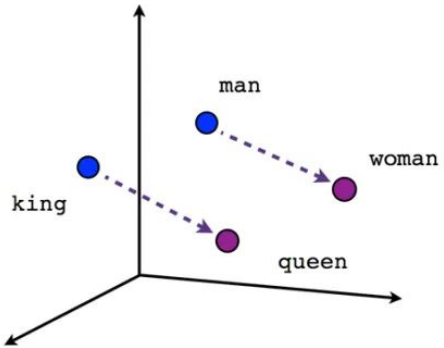
This doesn’t work well …

Why to use embeddings instead:

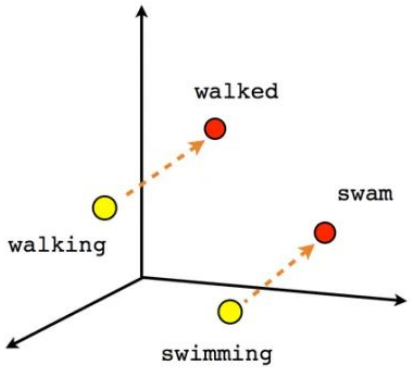
- Our encoded vectors do not adequately represent the data
  - The range of values for each columns is different, for example columns deal with account_balance skew the data
- Embeddings capture semantic relationships and context, which can enhance the quality of similarity measures.

Let’s use a neural network to create embeddings for each of our users!

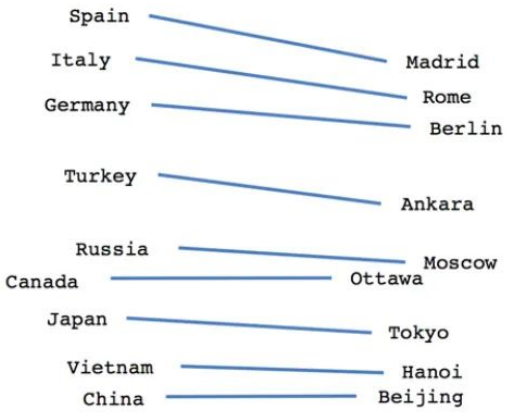
Intro to Embeddings



Male-Female

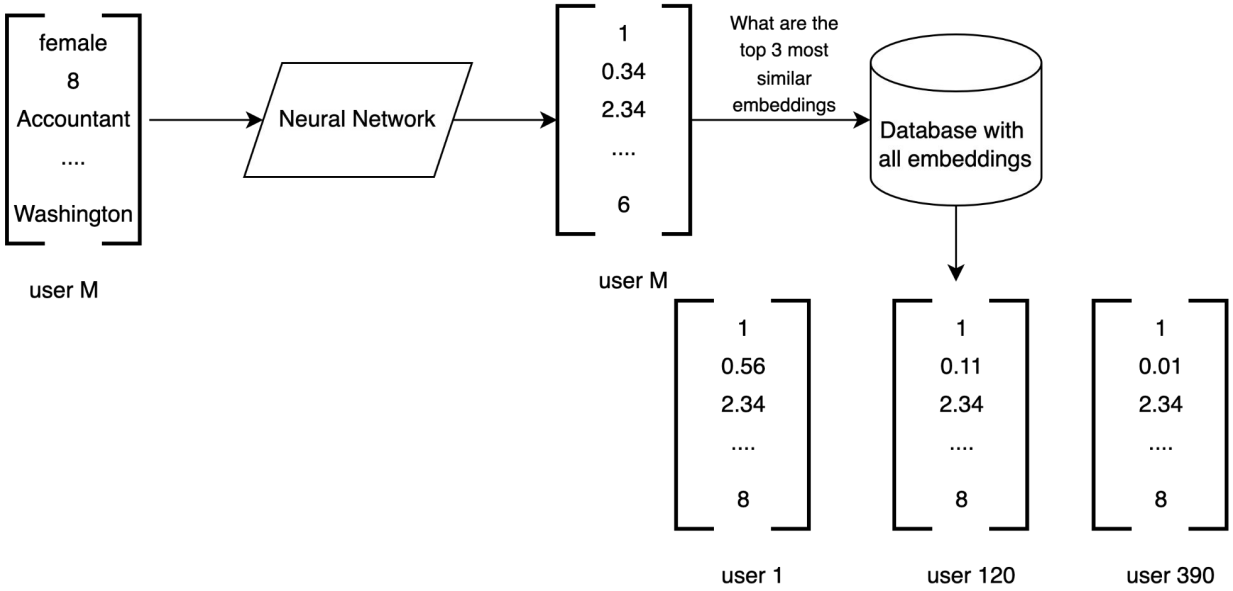


Verb tense



Country-Capital

Intro to Ranking using Machine Learning



Enhance Performance

- Using Prefect to run your predictions daily
- Using a vector feature store to support storing and computing vector similarities
- Tracking data drift and model metrics using Whylabs

ANITA B.ORG 2023
GRACE HOPPER CELEBRATION
THE WAY FORWARD

Questions?

/ANITA B.ORG 2023 GRACE HOPPER CELEBRATION | THE WAY FORWARD

Thank You

Feedback survey

julie-quintero

rashmi-raghunandan

anusha-jamkhandi