

Data Science Cup 2017 – Aufgabenstellung

Es gibt vier Aufgabenbereiche, die unterschiedliche Aspekte der Daten adressieren. In jedem Aufgabenbereich werden Punkte vergeben, die dann zu einer Gesamtwertung summiert werden. Die Aufgabenbereiche sind:

Aufgabe 1: Vorhersage Absatzzahlen

Gegeben sind jeweils die Top 5 der umsatzstärksten Items aus den Marktgruppen Ships, Ship Equipment, Ammunition & Charges und Manufacture & Research. Für jedes dieser 20 Items sollen auf verschiedenen zeitlichen Granularitäten und Horizonten Verkaufszahlen prognostiziert werden. Basierend auf historischen Daten auf Tagesgranularität bis zum 23.12.2015 sind dies: tägliche Verkäufe für die Tage vom 24.12.15 bis 01.01.16, wöchentliche Verkäufe für den Januar 2016, sowie die monatlichen Verkaufszahlen für Januar, Februar und März 2016. In Summe sind 16 (9+4+3) Prognosewerte pro Item zu berechnen.

Bewertung: Für jeden Prognosewert P wird die Abweichung vom tatsächlichen Wert A mittels des Symmetric mean absolute percentage error (SMAPE) gemessen. Alle Einzelfehler werden addiert und gemittelt um den Gesamtfehler für die Aufgabe zu bestimmen. Gewonnen hat die Einreichung mit dem geringsten Gesamtfehler. Der SMAPE ist definiert als:

$$SMAPE = \frac{|P - A|}{(|A| + |P|)/2}$$

und liefert Werte zwischen 0% und 200%.

Items:

Marktgruppe	Type ID	Item Name
Ships	29984	Tengu
	22548	Mackinaw
	12005	Ishtar
	22544	Hulk
	645	Dominix
Ship Equipment	15681	Caldari Navy Ballistic Control System
	11578	Covert Ops Cloaking Device II
	17559	Federation Navy Stasis Webifier
	2281	Adaptive Invulnerability Field II
	15806	Republic Fleet Gyrostabilizer
Ammunition & Charges	28668	Nanite Repair Paste
	30488	Sisters Core Scanner Probe
	32014	Navy Cap Booster 800
	2629	Scourge Fury Heavy Missile
	32006	Navy Cap Booster 400
Manufacture & Research	34	Tritanium
	36	Mexallon
	35	Pyerite
	38	Nocxium
	37	Isogen

Aufgabe 2: Vorhersage Preisentwicklung

Gegeben sind 3 Indizes, die verschiedene Teilmärkte abdecken. Diese Indizes liegen als Views vor und enthalten jeweils eine Zeitreihe für den durchschnittlichen Preis des Index auf Tagesgranularität. Basierend auf historischen Daten bis zum 23.12.2015 sind folgende Prognosen für den Preis zu berechnen: tägliche Werte für die Tage vom 24.12.15 bis 01.01.16, wöchentliche Werte für den Januar 2016, sowie monatliche Werte für Januar, Februar und März 2016. In Summe sind 16 (9+4+3) Prognosewerte pro Index zu berechnen.

Die 3 Indizes sind wie folgt definiert:

- Consumer Price Index (CPI): Ammunition, Consumer Products, Drones, Implants, Skills, Tech I Modules, Tech I ships, Tech II modules, Tech II ships
- Mineral Price Index (MPI): Ore, Minerals
- Secondary Producer Price Index (SPPI): Blueprints, Advanced Capital Components, Standard Capital Components, R.A.M., Advanced Components, Rigs

Bewertung: Jeder Prognosewert erhält eine Wertung zwischen 0 und 2 Punkten, die sich aus zwei Bestandteilen zusammensetzt. Die korrekte Vorhersage des Preistrends (steigt/fällt) wird mit einem Punkt bewertet. Der zweite Teil der Wertung ergibt sich über den SMAPE aus der Abweichung von Prognose- und Realwert. Die Wertung ist definiert als:

$$\text{Preistrend} + \frac{200\% - \text{SMAPE}}{200\%}$$

Durch die zwei Komponenten der Bewertung wird sichergestellt, dass eine Prognose mit falschem Preistrend und geringem SMAPE keine höhere Wertung als eine Prognose mit korrektem Preistrend erhalten kann. Gewonnen hat die Einreichung mit der höchsten Gesamtwertung.

Aufgabe 3: Vorhersage Marktanteile

Gegeben sind drei Marktsegmente mit dazugehörigen Produktgruppen. Für jede Produktgruppe liegen die Umsätze auf monatlicher Granularität vor (A3_queries.sql). Ziel dieser Aufgabe ist es, den prozentualen Anteil jeder Produktgruppe am Umsatz des jeweiligen Marktsegments vorherzusagen. Basierend auf historischen Daten bis zum 23.12.2015 sind die monatlichen Marktanteile für jede Produktgruppe für das erste Quartal 2016 zu prognostizieren.

Bewertung: Jeder Prognosewert erhält eine Wertung zwischen 0 und 2 Punkten, die sich aus zwei Bestandteilen zusammensetzt. Die korrekte Vorhersage des Wachstums (wächst/schrumpft) wird mit einem Punkt bewertet. Der zweite Teil der Wertung ergibt sich über den SMAPE aus der Abweichung von Prognose- und Realwert. Die Wertung ist definiert als:

$$\text{Wachstum} + \frac{200\% - \text{SMAPE}}{200\%}$$

Durch die zwei Komponenten der Bewertung wird sichergestellt, dass eine Prognose mit falschem Wachstum und geringem SMAPE keine höhere Wertung als eine Prognose mit korrekter Wachstumsprognose erhalten kann. Gewonnen hat die Einreichung mit der höchsten Gesamtwertung.

Marktsegmente:

Marktsegment	Produktgruppe	MarketgroupID
Ships		4
	Shuttles	391
	Frigates	1361
	Cruisers	1367
	Destroyers	1372
	Battlecruisers	1374
	Battleships	1376
	Capital Ships	1381
	Industrial Ships	1382
	Mining Barges	1384
	Special Edition Ships	1612
	Rookie Ships	1815
Missile Launchers		140
	Rocket Launchers	639
	Light Missile Launchers	640
	Rapid Light Missile Launchers	641
	Heavy Launchers	642
	Cruise Launchers	643
	Torpedo Launchers	644
	XL Launchers	777
	Heavy Assault Launchers	974
	Rapid Heavy Missile Launchers	1827
	Rapid Torpedo Launchers	2247
Materials		533
	Raw Materials	1031
	Gas Clouds Materials	1032
	Ice Products	1033
	Reaction Materials	1034
	Planetary Materials	1332
	Minerals	1857
	Salvage Materials	1861
	Faction Materials	1897

Aufgabe 4: Optimaler Verkaufsplan

Gegeben ist ein Warenbestand mit verschiedenen Items in unterschiedlichen Stückzahlen. Ziel dieser Aufgabe ist es, eine Menge von Verkaufsgeboten so zu definieren, dass der gesamte Verkaufserlös maximiert wird. Ein Verkaufsgebot wird definiert durch einen Ort (stationsid), ein Datum, ein Item, die zu verkaufende Stückzahl und einen Stückpreis. Es dürfen beliebig viele Verkaufsgebote an beliebig vielen Orten eingestellt werden. Verkaufsgebote sind nur an dem Tag und in der Station gültig für die sie definiert wurden. Verkäufe in anderen Stationen werden durch Angebote in regional nahegelegenen Stationen nicht beeinflusst. Gebote sind auch an durch Spieler einnehmbaren Stationen möglich (stationid beginnt mit 61..., siehe PDF "Lies Mich") und werden regulär bewertet. Da es keine effiziente Möglichkeit gibt, den Teilnehmern Rückmeldungen zu eventuellen Restbeständen zu geben, verfallen nicht verkaufte Einheiten und stehen für weitere Verkaufsgebote nicht mehr zur Verfügung. Die aufsummierte Stückzahl eines Items über alle Verkaufsgebote darf nicht höher als der Warenbestand des Items sein, d.h. zum Monatsbeginn ist der gesamte Warenbestand gegeben und es kommen keine weiteren Stückzahlen hinzu, Leerverkäufe sind nicht erlaubt. Der zeitliche Rahmen für alle Verkäufe ist der Januar 2016. Als Datengrundlage stehen die Stückzahlen und durchschnittlichen Preise aller Kauf- und Verkaufsgebote sowie der tatsächlichen Verkäufe pro Tag, Item und Station / System bis zum **23.12.2015** zur Verfügung.

Bewertung: Zur Bewertung der Verkaufsgebote werden die Einzeltransaktionen der jeweiligen Tage im Januar benutzt. Diese Daten wurden zur Berechnung der Tagesaggregate genutzt, liegen in der feingranularen Form aber nur dem ScaDS vor. Jede Einreichung wird einzeln im Markt bewertet, es besteht keine weitere zusätzliche Konkurrenz im Markt durch andere Teilnehmer. Die Bewertung eines Verkaufsgebotes wird wie folgt vorgenommen: Für das angegebene Item werden alle Verkaufstransaktionen für den spezifizierten Tag und Ort untersucht. Jeder tatsächliche Verkauf der zu einem höheren oder gleichen Preis als das zu bewertende Verkaufsgebot getätigt wurde, wird verdrängt und gutgeschrieben. Beispiel: Verkaufsgebot A (Item X, Tag Y, Ort Z, Stückzahl: 200, Preis: 5000) solle bewertet werden. Die tatsächlichen Verkäufe für Item X belaufen sich auf 400 Stück, wovon 150 für einen Betrag größer gleich 5000 verkauft wurden, während der Rest zu einem günstigeren Preis abgesetzt wurde. Diese 150 Einheiten werden durch Verkaufsgebot A bedient, da es günstiger als der tatsächliche Preis ist. Dadurch ergibt sich ein Erlös von 750k. Die restlichen 50 Einheiten von A werden aufgrund ihres, im Vergleich zu hohen, Preises nicht verkauft und verfallen. Gewonnen hat die Einreichung mit dem höchsten Gesamterlös.

Warenbestand:

Typeid	Item Name	Stückzahl
34	Tritanium	200.000.000
1230	Veldspar	40.000.000
43	Antibiotics	1.000.000
28668	Nanite Repair Paste	40.000
29984	Tengu	10

Ermittlung des Gesamtsiegers

Jede Aufgabe wird separat bewertet und die Einreichungen entsprechend der erreichten Punkte platziert. Die ersten drei Plätze bekommen dann Punkte für die Gesamtwertung zugewiesen: 1. Platz: 3 Punkte, 2. Platz: 2 Punkte, 3. Platz: 1 Punkt. Diese Punkte werden für die vier Aufgaben addiert, um die Gesamtwertung zu erhalten. Die Einreichung mit der höchsten Gesamtwertung gewinnt den Data Science Cup. Maximal können 12 Punkte erreicht werden.