CASE STUDY

NEVEŘEJNÁ

BIG DATA

SPOLUPRÁCE T-MOBILE A VŠE

RICHARD NOVÁK

ANTONÍN PAVLÍČEK

T-Mobile a spolupráce s VŠE

Team

- Big data agilní team
  - T-Mobile každý den cca 6-10 lidí
- Zapojení VŠE
  - Antonín Pavlíček, každé pondělí
  - 2 studenti
    - Zapojeni do výběru nástrojů

Technologie

- Zdroje dat
- Uložení dat
- Vizualizační a Prezentační vrstva

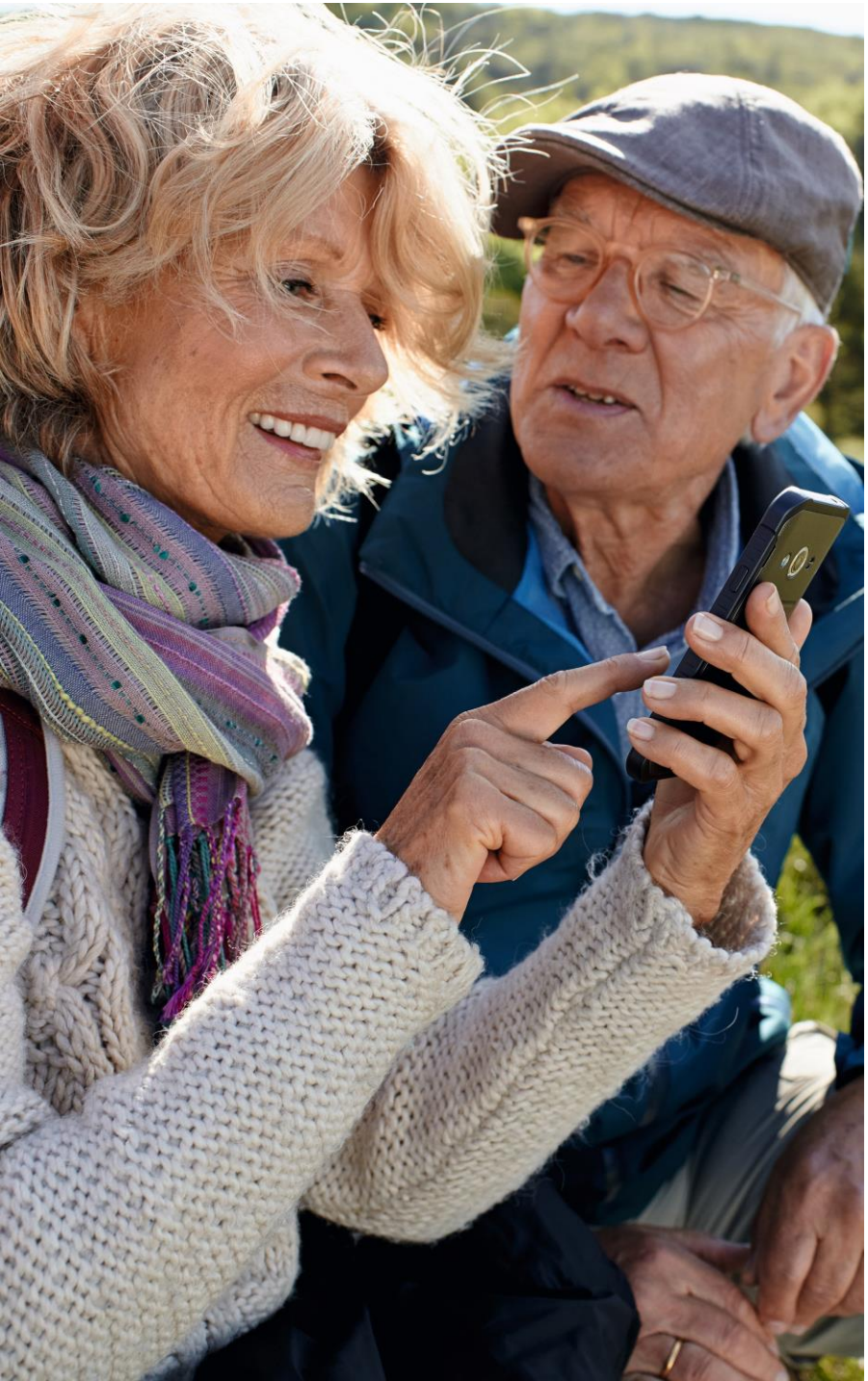
Big Data

- Geolokační úlohy a řešení
  - Např. Horská střediska
- Finanční instituce
  - Credit scoring
- Ostatní
  - Geomarketing.. atd

Případové studie

- Diplomová práce na výběr vizualizačních nástrojů
- Sociální sítě
- Oblast Etiky ve vztahu k Big Data

Další spolupráce



SOUKROMÍ A BEZPEČNOST
DATA JSOU ANONYMIZOVANÁ

Zásady zachování soukromí

- Data jsou bezpečně uchovávána a přístup k nim je striktně omezený
- Všechna data jsou před zpracováním anonymizovaná
- Data nemohou být zpětně přiřazena k uživateli
- Výstupy jsou vytvářeny v agregované podobě
- **Nesledujeme konkrétního zákazníka, ale chování celé populace**

Mobilní signalizační data

Anonymizace

Z pohledu legislativy je vždy před zpracováním nutné provést anonymizaci, která zabrání identifikaci jednotlivce.

Výstupní reporty jsou vždy v agregované podobě v podobě popisných metadat např. chování cílové skupiny (muži x ženy, Praha x horská střediska)

Zdroje data

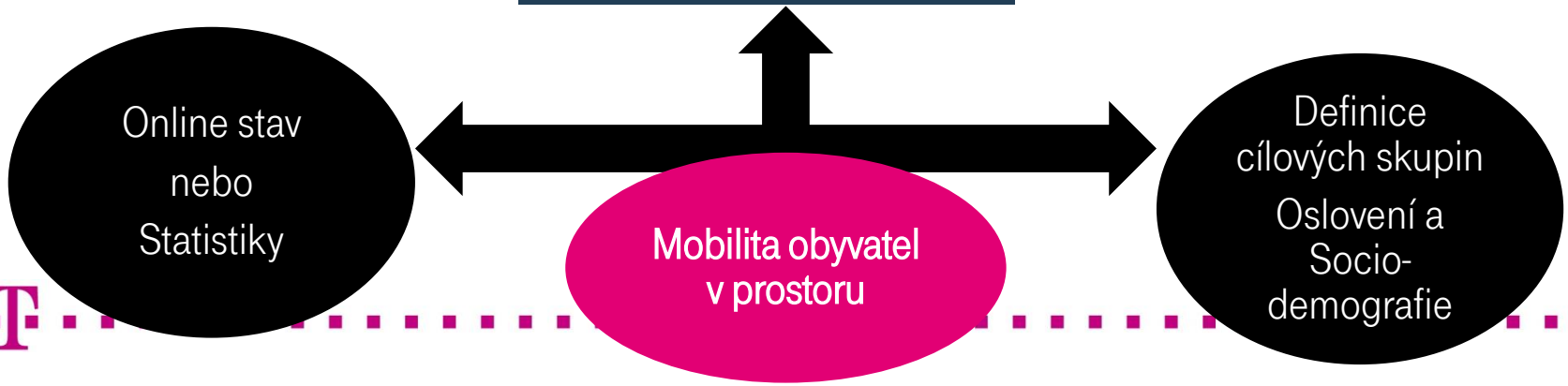
Zdrojem dat je mobilní signalizace vztažená k jednotlivé mobilní buňce.

EDR vzniká při hovoru, zaslaní SMS, datový přenos, převod mezi vysílači nebo tzv. intervalový „location update“

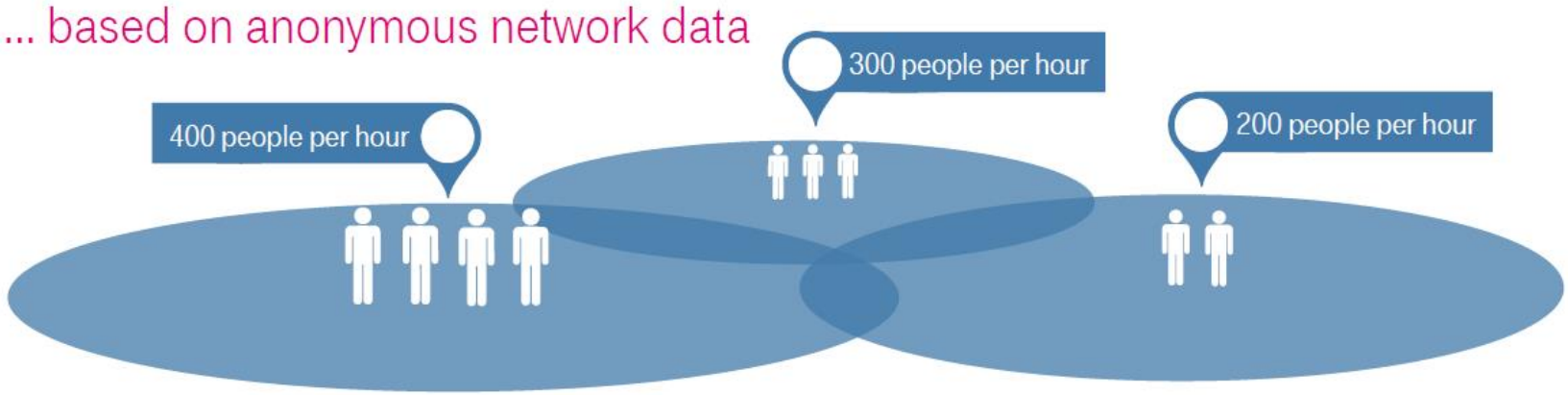
Zpracování

Na základě konkrétního USE CASE signalizační data dále obohacuje např. o CRM údaje a dopočítáváme žádoucí výstup.

Výstup může být: Znalost terminálů v dané lokalitě, matice zdrojů a cílů, zohlednění časový řezů, extrapolace vzorku na celé obyvatelstvo ..atd



Mobilita obyvatel v prostoru



Socio-demografie a CRM Data:

USE CASES
Horská střediska

ZADÁNÍ
CO JSME SLEDOVALI

Cílem analýzy je pomocí anonymních dat o pohybu našich zákazníků z mobilní sítě změřit návštěvnost horských středisek

Sledované období:

- 29. 1. – 11. 2. 2016

Zaměřili jsme se na česká horská střediska:

- Harrachov (Krkonoše)
- Kohútka (Javorníky)
- Lipno (Šumava)
- Pec pod Sněžkou (Krkonoše)
- Rokytnice nad Jizerou (Krkonoše)
- Špindlerův Mlýn (Krkonoše)

Přesnější vymezení
JAKÁ DATA JSME POUŽILI K ANALÝZE

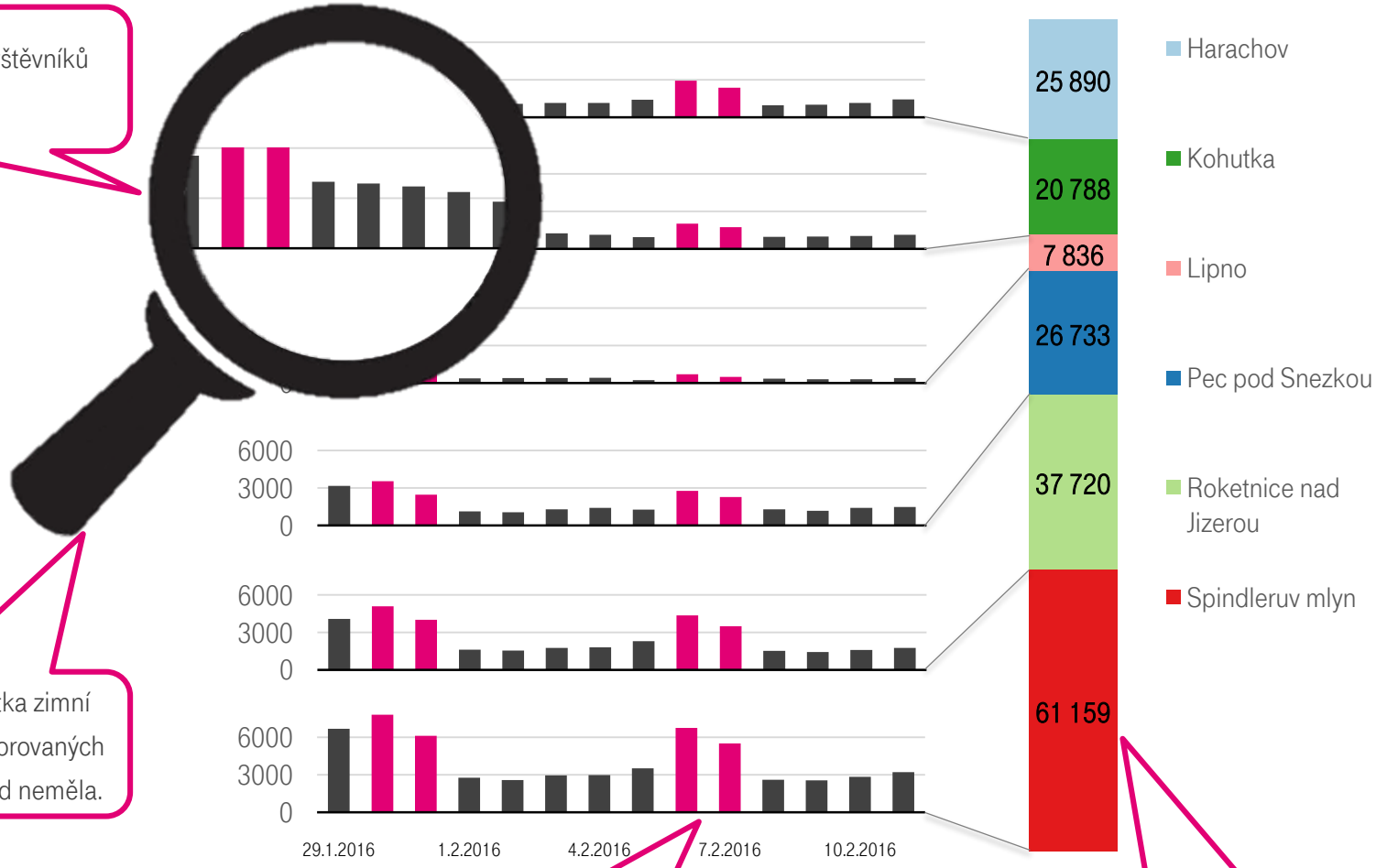
- **Zdroje dat**
  - Mobilní signalizace
  - Sociodemografická data ze CRM

- **Sledované atributy:**
  - Navštívené horské středisko
  - Doba návštěvy
  - Země původu a u zákazníků TMCZ lokalitu bydliště
  - Pohlaví a věk
  - Počet návštěvníků v jednotlivých dnech

- Výsledné hodnoty vztahujeme k počtu unikátních návštěvníků

POROVNÁNÍ NÁVŠTĚVNOSTI

Nejmenší výkyvy v počtu návštěvníků má areál Kohútka.



3. února navštívil areál Kohútka zimní inspektor televize Nova, v pozorovaných datech však akce žádný dopad neměla.

... nejsilnějšími dny jsou soboty a neděle, z pracovních dní pátky.

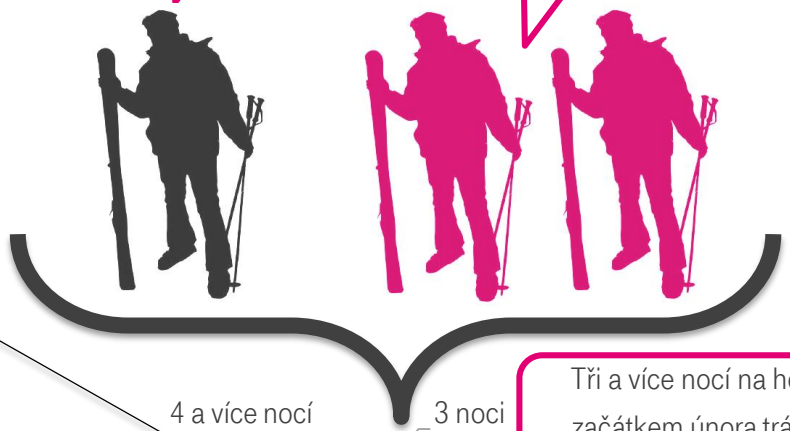
V počtu unikátních návštěvníků zvítězil Špindlerův Mlýn, měl o polovinu více, než druhá Rokytnice nad Jizerou, ...

– Inte

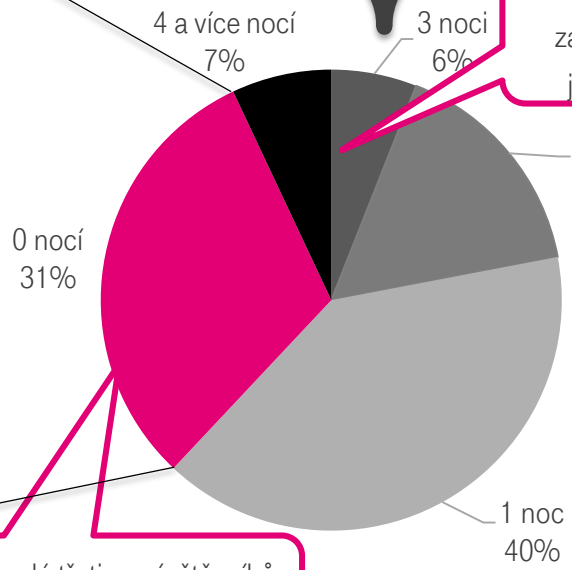
JAK DLOUHO NÁVŠTĚVNÍCI ZŮSTÁVAJÍ

Za každého návštěvníka v průběhu pracovního týdne ...

.. přijedou o víkendu dva!



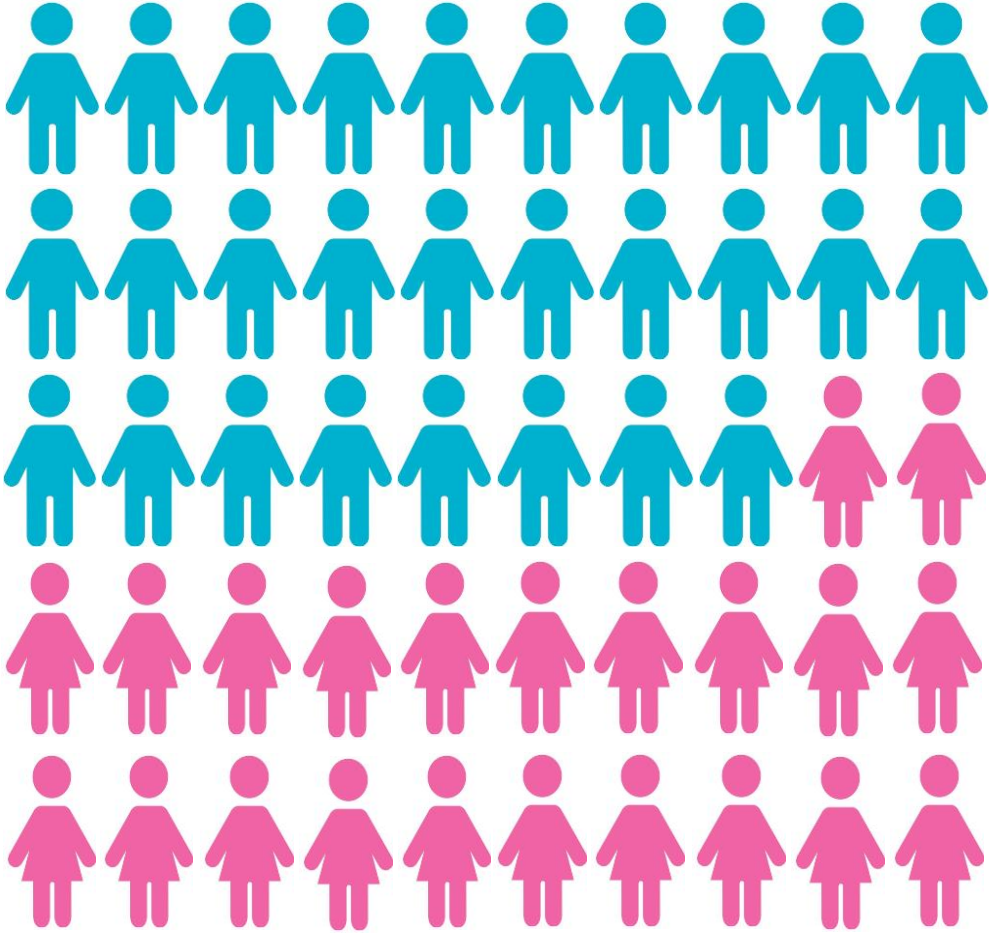
... z toho 90 % nelyžuje celý den a odjede dříve než po 8 hodinách



Tři a více nocí na horách začátkem února tráví jen jeden z 8 návštěvníků.

Necelá třetina návštěvníků je v areálu je jeden den ...

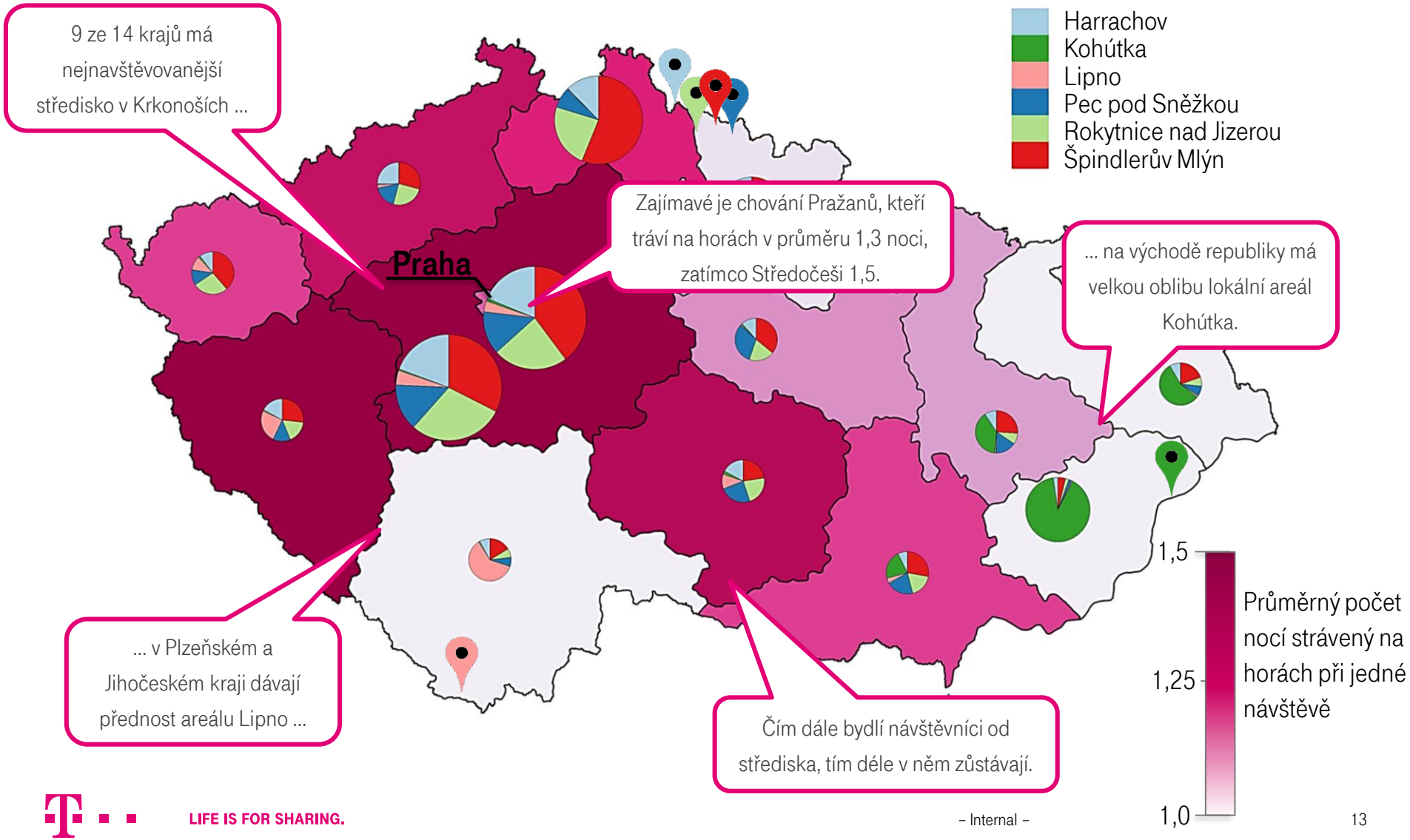
… NEJVĚTŠÍ PROCENTO MUŽŮ JE V HARRACHOVĚ …



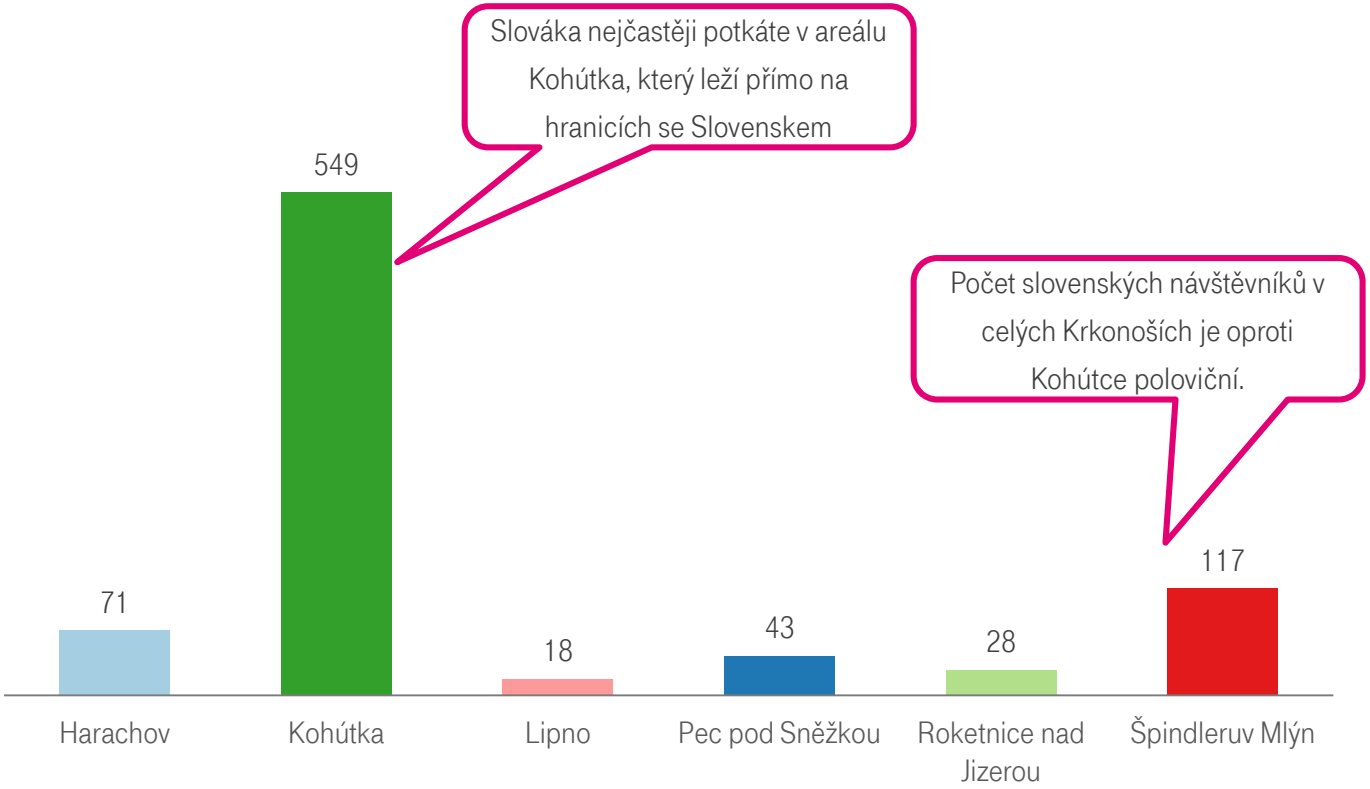
56 %

44 %

KDE NÁVŠTĚVNÍCI BYDLÍ A KAM JEZDÍ



KAM JEZDÍ SLOVÁCI?



USE CASE – BANK B
CREDIT SCORING

EXECUTIVE SUMMARY

OBJECTIVES

- To improve loan approval scoring for bank
- To use credit scoring from 3rd party (T-Mobile CZ) for the purpose of the retail bank
- Use especially for clients unknown to bank so far

PARTNERS DATA MODEL

- BANK B, T-Mobile CZ
- Data provided by both partners are hashed
- Two years of cash loan application data (BANK B) and both customers behavior for the same period (TMCZ)

BANK SEGMENTS

EXI = existing clients
NTB = new-to-bank
NTM = new-to-market

RESULT OF THE PROJECT

- Third party developed a model based on encrypted data
- T-Mobile CZ implements the model
- T-Mobile CZ and BANK B start a two-side communication (e.g. web service)

USE CASES
T-MOBILE CZ CASES FOR FINANCIAL SECTOR

Credit Scoring	▪ T-Mobile provides scoring based on behavioral data ▪ We covers even New-To-Market and New-To-Bank customers
Churn prediction	▪ Customer behavior on the internet, e.g. competition web site visit ▪ SMS/MMS marketing and retention campaign targeting
Fraud prediction	▪ Customer background check based on T-Mobile data ▪ Suspicious customer internet behavior analysis
Branch network	▪ Optimal branch placement in relation to customer distribution ▪ Locality preselection for new branch opening
Customer targeting	▪ Based on customers current locality (near to branch, airport, etc.) ▪ Product offers based on the internet activity and sociodemographic profile

Backup